POSTTRANSCRIPTIONAL GENE REGULATION

BY THE *DMD* 3'UTR

by

Charles Aaron Larsen

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Human Genetics

The University of Utah

August 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Charles Aaron Larsen**

has been approved by the following supervisory committee members:

Michael Howard	, Chair	**05/28/2014** Date Approved
Brenda Bass	, Member	**05/28/2014** Date Approved
Dana Carroll	, Member	**05/28/2014** Date Approved
Gabrielle Kardon	, Member	**05/28/2014** Date Approved
Robert Weiss	, Member	**05/28/2014** Date Approved

and by **Lynn Jorde** , Chair/Dean of

the Department/College/School of **Human Genetics**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Duchenne muscular dystrophy (DMD) and the milder Becker muscular dystrophy (BMD) are both caused by mutations in the *DMD* gene, which encodes the protein dystrophin. Its regulation is of therapeutic interest as even small changes in expression of functional dystrophin can significantly impact disease severity in humans. While tissue specific distribution and transcriptional regulation of dystrophin has been characterized, the posttranscriptional regulation of dystrophin is not well understood. 3' UTRs have been shown to regulate gene expression posttranscriptionally by altering mRNA stability, localizing mRNA, or directly affecting translation, and the *DMD* 3' UTR is particularly interesting in that it contains very large, highly conserved regions that are more conserved than the coding region of the gene.

In this dissertation, we take advantage of C2C12 mouse muscle cells and the pHRL Renilla reporter construct to dissect the regions of the *DMD* 3' UTR responsible for regulating expression during myogenesis. We identify highly conserved regions in the 3' UTR that alter mRNA stability and translation of a reporter construct during differentiation. We characterize the *DMD* 3' UTR variants in a population of 1,222 humans that include DMD and BMD patients. We conclude that if disease causing mutations exist in the *DMD* 3' UTR they are either extremely rare or cause a disease phenotype in nonmuscular tissues, but that variation in the 3' UTR could alter disease phenotype when a second mutation is present or affect the efficacy of developing

treatments in human patients.

Canine models of DMD are important tools for developing treatments for human DMD patients because they better recapitulate the disease phenotype and have an immune system more similar to humans' than other model organisms. In this dissertation, we characterize three canine models of DMD, each containing different classes of mutations that can be used to facilitate preclinical studies directed toward these specific mutation classes in developing therapeutic approaches.

This dissertation adds to the understanding of how dystrophin is regulated, the role the *DMD* 3' UTR has in dystrophin regulation and pathogenesis of disease, and characterizes additional canine models of DMD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Chapters 1 and 2 of this dissertation represent original work conducted under the supervision and guidance of Dr. Michael T. Howard, PhD (University of Utah, Human Genetics Department). A manuscript containing the contents of Chapter 1 of this dissertation has been accepted for publication in the journal *Neuromuscular Disorders* with myself and Dr. Michael T. Howard (Human Genetics Department, University of Utah) as coauthors. Chapter 2 represents original, unpublished work. Chapter 3 of this dissertation represents original work conducted under the supervision and guidance of Dr. Kevin Flanigan (Nationwide Children's Hospital, Columbus, Ohio) during his tenure at the University of Utah Human Genetics Department. The canine mutations described in Chapter 3 are found in Table 1 in the manuscript "Canine Models of Duchenne Muscular Dystrophy and Their Use in Therapeutic Strategies" by Kornegay et al. published in Mammalian Genome in February, 2012, and referenced as 'Larsen CA et al. unpublished' of which I am a co-author. In addition to the work presented in this dissertation, I contributed the work for Figure 1.b in the manuscript "Becker Muscular Dystrophy Due to an Inversion of Exons 23 and 24 of the DMD gene" by Flanigan et al., published in Muscle Nerve in November of 2011, of which I am also a co-author.

I give special thanks to Dr. Michael T. Howard for his immense support and guidance in completing this dissertation. I thank Dr. Robert B. Weiss (University of

Utah, Department of Human Genetics) for providing human sequencing data used in this dissertation, and for his comments and suggestions throughout the course of this work. I thank Dr. Brenda L. Bass (University of Utah, Departments of Biochemistry and Human Genetics), Dr. Dana Carroll (University of Utah, Department of Biochemistry), and Dr. Gabrielle Kardon (University of Utah, Department of Human Genetics) for their comments and suggestions on the work contained in this dissertation. I would also like to thank Norma M. Wills (University of Utah, Department of Human Genetics), Chris Anderson (University of Utah, Department of Human Genetics), and Laura Taylor (Nationwide Children's Hospital, Columbus, Ohio) for technical assistance provided throughout the course of this work.

INTRODUCTION

Muscular dystrophies are a group of disorders characterized by progressive muscle degeneration. The most common and most severe form is the X-linked, recessive Duchenne muscular dystrophy (DMD). DMD was first described in 1851 by Edward Meryon and published the following year (Emery, 2002; Meryon, 1852), but became associated with the French neurologist Guillaume Duchenne who published the clinical and histological features of the disease several years later (Duchenne, 1867; Emery, 2002). Although patients afflicted with DMD are asymptomatic at birth, the effects of progressive muscle degeneration begin to manifest in early childhood. Patients often show difficulty walking, running, and climbing stairs by five years of age (Blake et al., 2002; Emery, 2002; Sussman, 2002). Weakness of the pelvic girdle and lower limb muscles results in the Gower's maneuver, where a child must climb up his thighs to stand up from a lying down position. As the degenerative process continues, muscle tissue is replaced by a fibrous connective tissue and the muscles become enlarged, an observation often seen in patients at three to four years old that present with enlarged calf muscles. As the lower limb muscles continue to weaken, the patient loses the ability to walk and becomes wheelchair bound, typically by 12 years of age. Ultimately, the intercostals and cardiac muscles begin to be affected and death occurs as early as the late teens or early 20s. However, with proper management and modern techniques, such as assisted ventilation, life can be extended into the late 20s and beyond (Eagle et. al., 2007;

Simonds et al.,1998; Wagner et al., 2007).

The milder Becker muscular dystrophy (BMD) was first described in the 1950s (Becker, 1962; Becker and Kiener, 1955) and was found to cause muscle degeneration and weakness similar to Duchenne muscular dystrophy but at a much slower rate. In BMD patients, onset of clinical symptoms can begin around 12 years of age, but some patients do not show symptoms until much later in life. Loss of ambulation in BMD patients is highly variable and can occur as early as adolescence or well into late adulthood. BMD patients typically survive into the fourth or fifth decades of life.

Cognitive defects are seen in both DMD and BMD patients. Approximately one third of DMD patients have cognitive impairments with the DMD population having a mean full scale intelligence quotient (IQ) score one standard deviation lower than the population average (Cotton et al., 2005). Neurological disorders, such as autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD), also occur at an increased frequency in DMD and BMD patients (Hendriksen and Vles, 2008; Wu et al., 2005).

Both DMD and BMD are caused by mutations in the *DMD* gene (Koenig et al., 1987; Muntoni et al., 2003). The DMD gene is the largest known human gene, spanning more than 2 megabases and comprising 79 exons on the X chromosome. Tissue specific promoters within the *DMD* gene drive the expression of several isoforms of the protein dystrophin (Hoffman et al., 1987; Muntoni et al., 2003). Three full length isoforms exist, each having a unique first exon spliced to the remaining 78 exons of the *DMD* gene and include the isoforms Dp427m, a muscle specific isoform (Koenig et al., 1987; Monaco et al., 1986); Dp427c, expressed primarily in cortical neurons in the brain (Nudel et al.,

1989); and Dp427p, expressed primarily in cerebellar purkinje cells (Gorecki et al., 1992; Holder et al., 1996). The expression of several truncated dystrophin isoforms is driven by promoters that lie within the introns of the *DMD* gene and are coded by the exons downstream of the promoter. These include Dp260, a truncated dystrophin isoform expressed in the retina (D'Souza et al., 1995); Dp140, expressed in the central nervous system and kidneys (Lidov and Kunkel, 1997; Lidov et al., 1995); Dp116, expressed in Schwann cells (Byers et al., 1993); and Dp71, which is ubiquitously expressed (Austin et al., 1995; Lederfein et al., 1992). All of these *DMD* mRNA isoforms encode the C-terminal domain of dystrophin and are thought to share the same 3' UTR.

The most extensively studied dystrophin isoform is the muscle specific isoform, Dp427m. It is a large, 427 kDa protein that localizes to the sarcolemma and acts as a scaffold for a complex of proteins including the dystroglycans, sarcospan, syntrophin, and dystrobrevin to form the dystrophin associated glycoprotein complex (DGC). The DGC anchors the cytoskeleton of muscle fibers to the extracellular matrix and has been shown to play an important structural role in maintaining muscle fiber integrity. The dystrophin protein can be organized into three distinct domains; the N-terminal domain that contains actin binding domains that bind to the cytoskeleton inside the cell, a C-terminal domain that interacts with proteins of the DGC at the subsarcolemmal space, and a large rod domain found in the middle of the protein that connects the N- and C-terminal domains and is thought to provide dystrophin with the flexibility it needs to function during muscle contractions. In addition to the structural role, dystrophin has also been shown to mediate signal transduction cascades through the C-terminal domain (Blake et al., 2002; Cohn and Campbell, 2000), and to regulate miRNAs in muscle cells

(Cacchiarelli et al., 2010; Marrone and Shcherbata, 2011).

Other dystrophin isoforms have important functions in nonmuscular tissues. In the brain, the dystrophin glycoprotein complex is involved in several cellular processes in brain development and function (Waite et al., 2012; Waite et al., 2009). In addition to the full length brain isoforms (Dp427c and Dp427p), the shorter isoforms Dp71 and Dp140 are also expressed in the brain. Dp71 has been shown to form dystrophin associated protein complexes at the plasma membrane and nucleus and is the most abundant isoform in the brain (Tadayoni et al., 2012). Dp140 is developmentally regulated in the brain with higher expression during fetal development compared to adults (Lidov et al., 1995). There is a correlation with mutations at the distal end of the *DMD* gene and cognitive impairment in human patients, and mutations affecting these dystrophin isoforms could explain the neurological disorders seen in DMD and BMD patients. The dystrophin isoform expressed in the retina, Dp260, is involved in synaptic maturation and attachment of the retina to the vitreous (Rodius et al., 1997; Schmitz and Drenckhahn, 1997), and ophthalmological defects, such as red green vision impairments, have been observed at a higher frequency in DMD and BMD patients with mutations downstream of exon 30 that lie within this isoform (Costa et al., 2007).

Whether a patient develops DMD or BMD is largely dependent on the type of mutation found in the *DMD* gene. The more severe DMD phenotype is typically caused by mutations that severely reduce or eliminate synthesis of functional dystrophin protein, such as those that disrupt the reading frame (Flanigan et al., 2009; Flanigan et al., 2011; Monaco et al., 1988; Muntoni et al., 2003), whereas the less severe BMD is caused by mutations that maintain the open reading frame and allow for residual expression of

functional dystrophin protein (Flanigan et al., 2009; Flanigan et al., 2011; Monaco et al., 1988; Muntoni et al., 2003). This observation has been described as the reading frame rule (Monaco et al., 1988), and holds true even in some very drastic cases. For example, one BMD patient was found to have a very large in frame deletion in the rod domain that deleted 46% of the coding region of the *DMD* gene (England et al., 1990). Yet this patient had very mild muscular dystrophy and was still ambulant at age 61 (England et al., 1990). Observations like this inspired the theory that DMD patients could be treated if partially functional dystrophin expression could be restored, and led to the development of several therapeutic approaches to treat DMD patients (Fairclough et al., 2013; Muir and Chamberlain, 2009; Pichavant et al., 2011).

One such approach uses antisense oligonucleotides that bind to exon splicing junctions to skip exons in the final mRNA product. When out-of-frame mutations are present, the reading frame of the *DMD* transcript can be restored by skipping the exons containing and/or surrounding the mutation. This could potentially lead to the expression of a partially functional dystrophin protein and result in a BMD-like phenotype. It is estimated that 83% of all *DMD* mutations could be corrected by skipping specific exons (Aartsma-Rus et al., 2009; Muntoni and Wood, 2011). This technique has been successful in ameliorating the DMD phenotype in both animal models and human clinical trials product (Cirak et al., 2011; Goemans et al., 2011; Gurvich et al., 2008; Kinali et al., 2009; Lu et al., 2005; Mendell et al., 2013; van Deutekom et al., 2007).

Read through of premature stop codons is another approach to treating DMD patients (Barton-Davis et al., 1999; Finkel, 2010; Howard et al., 2004; Malik et al., 2010). Mutations that cause premature stop codons occur in ~15% of DMD patients (Aartsma-

Rus et al., 2006), and could be treated with drugs that promote readthrough. Initial human trials of one such drug, Ataluran, was shown to increase dystrophin expression in most patients and treated patients could walk further in a six-minute walk test compared to untreated patients (Finkel, 2010; McDonald et al., 2010). However, there was no correlation between dystrophin levels and the patients that could walk the furthest in this study. Other drugs with greater effectiveness have shown success in animal models (Kayali et al., 2012), but have yet to be tested in humans.

Delivery and expression of functional dystrophin using viral gene therapy in DMD patients is also being pursued (Fabb et al., 2002; Foster et al., 2012; Foster et al., 2008; Koo et al., 2011a; Koo et al., 2011b; Rodino-Klapac et al., 2010; Wang et al., 2000; Wang et al., 2007; Wang et al., 2012). However, the large 14 kilobase (kb) *DMD* mRNA cannot fit in the viral vectors which has been a major challenge with this approach. To overcome this, minidystrophin gene constructs (~3-6 kb) lacking the most nonessential *DMD* coding exons that express functional dystrophin protein have been designed, and have been successful in animal models of DMD (Fabb et al., 2002; Foster et al., 2012; Foster et al., 2008; Koo et al., 2011a; Koo et al., 2011b; Rodino-Klapac et al., 2010; Wang et al., 2000; Wang et al., 2007; Wang et al., 2012).

The fourth approach to treating DMD patients is by overexpressing utrophin. Utrophin is a dystrophin related protein with structural and functional similarities that has been shown to functionally compensate for the lack of dystrophin (Khurana and Davies, 2003; Miura and Jasmin, 2006). Utrophin levels have been increased by several methods and have shown success in animal models lacking dystrophin (Chakkalakal et al., 2008; Moorwood et al., 2013; Sonnemann et al., 2009; Tinsley et al., 2011). However, utrophin

is overexpressed in DMD patients (Mizuno et al., 1993) and the benefit of further increasing utrophin levels in humans has yet to be shown.

DMD homologues have been identified in several animal species from mammals (Bulfield et al., 1984; Carpenter et al., 1989; Cooper et al., 1988; Valentine et al., 1988), to fish (Bassett and Currie, 2004) and nematodes (Bessou et al., 1998), and have been very useful tools in understanding DMD pathogenesis and developing treatments for human patients. The best characterized animal model of DMD is the *mdx* mouse (C57BL/10ScSn-Dmd^{mdx}/J) that contains a nonsense mutation in exon 23 of the *DMD* gene and does not express dystrophin (Bulfield et al., 1984; Sicinski et al., 1989). However, the severe muscle degeneration and early death seen in human DMD patients is not seen in these mice (Collins and Morgan, 2003; Dangain and Vrbova, 1984). Instead, a relatively mild skeletal phenotype is observed. This is partly due to compensation for the lack of dystrophin by the functionally related protein utrophin, and although a dystrophin/utrophin double knockout mouse (*mdx/utrp-/-*) better mimics human DMD pathology (Janssen et al., 2005), it is not a perfect genetic model for DMD.

Larger mammals, such as the dog, better recapitulate the phenotype of human DMD. Canine X-linked muscular dystrophy (CXMD) dogs have an absence of dystrophin expression, and similar clinical features and disease progression as human DMD patients (Valentine et al., 1988). In addition, dogs have an immune system very similar to humans' (Wang et al., 2007), making the *cxmd* dog important for testing side effects of therapies, such as viral gene delivery, where the immune response in humans has been problematic (Cavazzana-Calvo et al., 2005; Mendell et al., 2012). The best characterized dog model to date is the Golden Retriever Muscular Dystrophy (GRMD)

dog (Cooper et al., 1988; Kornegay et al., 1988; Valentine et al., 1988). The GRMD dog has a point mutation in the splice acceptor site of intron 6, causing exon 7 to be skipped during splicing and results in a premature stop codon (Sharp et al., 1992).

Studies in the *mdx* mouse have shown that expression of even small amounts of dystrophin can lessen disease severity (van Putten et al., 2012; van Putten et al., 2013). Human clinical trials using developing therapies, such as exon skipping, have shown that these strategies can restore dystrophin expression in the majority of muscle fibers in treated muscle (Goemans et al., 2011; Kinali et al., 2009; Mendell et al., 2013; van Deutekom et al., 2007). However, dystrophin expression is highly variable between patients and typically results in less than 20% dystrophin levels compared to normal controls (Cirak et al., 2011; Goemans et al., 2011; Kinali et al., 2009; van Deutekom et al., 2007). Regulatory factors that affect dystrophin expression could account for some of the variability seen, and these factors could impact disease pathogenesis and effectiveness of developing therapies. While tissue specific distribution and transcriptional regulation of *DMD* have been well characterized, the posttranscriptional regulation of dystrophin synthesis is not well understood. The untranslated regions of genes, such as the 3' UTR, have been shown to regulate gene expression posttranscriptionally by altering mRNA stability, localizing mRNA, or directly affecting translation (Andreassi and Riccio, 2009; Chatterjee and Pal, 2009; Gramolini et al., 2001; Matoulkova et al., 2012; Mazumder et al., 2003; Pickering and Willis, 2005), and mutations in these regions can be pathogenic (Chatterjee and Pal, 2009). In one study, it was shown that *DMD* transcript stability was a bigger determinant of dystrophin protein levels in BMD patients than transcription rate, and increasing stability could improve the outcome of therapies such as exon skipping

(Spitali et al., 2013).

The *DMD* 3' UTR contains some particularly interesting characteristics that suggest it could be involved in posttranscriptional regulation of dystrophin. The 2.7 kb *DMD* 3' UTR is larger than the average 3' UTR length of ~1020 nt (Mignone et al., 2002) in human transcripts, and contains several large, highly conserved regions that could be regulating expression. The high amount of conservation in the *DMD* 3' UTR was first described over two decades ago comparing human and chicken sequences (Lemaire et al., 1988). Two large, conserved regions were identified, named Lemaire A and Lemaire D, both of which are more highly conserved than the protein coding region (Greener et al., 2002; Lemaire et al., 1988). Evidence for two potential regulatory elements within these regions of the *DMD* 3' UTR have been previously reported. miRNA-31 is predicted to bind to a site within Lemaire A, and is thought to suppress dystrophin synthesis in myoblasts prior to differentiation and in dystrophic muscle (Cacchiarelli et al., 2011). *In vitro* experimental evidence has also been presented for binding of the RNA binding protein vigilin to the Lemaire D region of the *DMD* 3' UTR (Kanamori et al., 1998). However, the biological significance of this binding with regard to dystrophin expression has not been explored.

To date, no well defined, disease causing mutations have been identified in the *DMD* 3' UTR in human DMD or BMD patients. There are two examples of BMD patients (Greener et al., 2002; McCabe et al., 1989) and two examples of DMD patients (Pillers et al., 1990; Todorova et al., 2008) with deletions that span the *DMD* 3' UTR. However, these deletions extend into the coding region of the *DMD* gene, delete other neighboring genes, or have not been fully analyzed to determine the extent of the

deletion, and disease causing mutations shown to be contained solely within the 3' UTR have yet to be identified. However, there is a small group of human patients (0.2% in one large study) (Flanigan et al., 2009) where dystrophin expression is absent, but they have no mutations in the coding region of the *DMD* gene, suggesting the possibility of disease causing mutations outside of the coding region, such as the *DMD* 3' UTR, that may have been overlooked by standard mutation screens.

The experiments discussed in this dissertation explore the role the *DMD* 3' UTR has in dystrophin expression in muscle tissue. The results presented here demonstrate that a highly conserved region overlapping Lemaire A at the beginning of the 3' UTR increases both mRNA stability and translation efficiency, whereas a highly conserved region overlapping Lemaire D at the distal end of the 3' UTR primarily affects mRNA abundance. We show evidence that the highly conserved element contains multiple regulatory elements and interacts with RNA binding proteins. We characterize the variants seen in *DMD* 3' UTRs from a population of 1,222 humans that include DMD and BMD patients, and conclude that a disease causing 3' UTR mutation is unlikely to exist in this group of patients. However, we could not rule out the possibility that variants in the 3' UTR could modify the disease phenotype when a second mutation is present or cause a disease phenotype in nonmuscular tissues. This dissertation concludes with a characterization of four canine models of DMD who presented in clinical veterinary practice with symptoms of weakness along with altered or absent dystrophin expression on muscle biopsy, and identify the mutations responsible in three such dogs. We developed genotyping assays for each mutation to aid in establishing research colonies of these dogs. Identification of these novel mutations adds to the catalog of

available dog models, and may facilitate developing therapeutic approaches directed toward these specific mutation classes in an animal model that more faithfully recapitulates the human disease.

References

Aartsma-Rus, A., Van Deutekom, J.C., Fokkema, I.F., Van Ommen, G.J., and Den Dunnen, J.T. (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve *34*, 135-144.

Aartsma-Rus, A., Fokkema, I., Verschuuren, J., Ginjaar, I., van Deutekom, J., van Ommen, G.J., and den Dunnen, J.T. (2009). Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations. Hum Mutat *30*, 293-299.

Andreassi, C., and Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. Trends Cell Biol *19*, 465-474.

Austin, R.C., Howard, P.L., D'Souza, V.N., Klamut, H.J., and Ray, P.N. (1995). Cloning and characterization of alternatively spliced isoforms of Dp71. Hum Mol Genet *4*, 1475-1483.

Barton-Davis, E.R., Cordier, L., Shoturma, D.I., Leland, S.E., and Sweeney, H.L. (1999). Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of mdx mice. J Clin Invest *104*, 375-381.

Bassett, D., and Currie, P.D. (2004). Identification of a zebrafish model of muscular dystrophy. Clin Exp Pharmacol Physiol *31*, 537-540.

Becker, P.E., and Kiener, F. (1955). [A new x-chromosomal muscular dystrophy]. Arch Psychiatr Nervenkr Z Gesamte Neurol Psychiatr *193*, 427-448.

Becker, P.E. (1962). Two families of benign sex-linked recessive muscular dystrophy. Rev Can Biol *21*, 551-566.

Bessou, C., Giugia, J.B., Franks, C.J., Holden-Dye, L., and Segalat, L. (1998). Mutations in the Caenorhabditis elegans dystrophin-like gene dys-1 lead to hyperactivity and suggest a link with cholinergic transmission. Neurogenetics *2*, 61-72.

Blake, D.J., Weir, A., Newey, S.E., and Davies, K.E. (2002). Function and genetics of dystrophin and dystrophin-related proteins in muscle. Physiol Rev *82*, 291-329.

Bulfield, G., Siller, W.G., Wight, P.A., and Moore, K.J. (1984). X chromosome-linked

muscular dystrophy (mdx) in the mouse. Proc Natl Acad Sci U S A *81*, 1189-1192.

Byers, T.J., Lidov, H.G., and Kunkel, L.M. (1993). An alternative dystrophin transcript specific to peripheral nerve. Nat Genet *4*, 77-81.

Cacchiarelli, D., Martone, J., Girardi, E., Cesana, M., Incitti, T., Morlando, M., Nicoletti, C., Santini, T., Sthandier, O., Barberi, L.*, et al.* (2010). MicroRNAs involved in molecular circuitries relevant for the Duchenne muscular dystrophy pathogenesis are controlled by the dystrophin/nNOS pathway. Cell Metab *12*, 341-351.

Cacchiarelli, D., Incitti, T., Martone, J., Cesana, M., Cazzella, V., Santini, T., Sthandier, O., and Bozzoni, I. (2011). miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy. EMBO Rep *12*, 136-141.

Carpenter, J.L., Hoffman, E.P., Romanul, F.C., Kunkel, L.M., Rosales, R.K., Ma, N.S., Dasbach, J.J., Rae, J.F., Moore, F.M., McAfee, M.B.*, et al.* (1989). Feline muscular dystrophy with dystrophin deficiency. Am J Pathol *135*, 909-919.

Cavazzana-Calvo, M., Lagresle, C., Hacein-Bey-Abina, S., and Fischer, A. (2005). Gene therapy for severe combined immunodeficiency. Annu Rev Med *56*, 585-602.

Chakkalakal, J.V., Miura, P., Belanger, G., Michel, R.N., and Jasmin, B.J. (2008). Modulation of utrophin A mRNA stability in fast versus slow muscles via an AU-rich element and calcineurin signaling. Nucleic Acids Res *36*, 826-838.

Chatterjee, S., and Pal, J.K. (2009). Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol Cell *101*, 251-262.

Cirak, S., Arechavala-Gomeza, V., Guglieri, M., Feng, L., Torelli, S., Anthony, K., Abbs, S., Garralda, M.E., Bourke, J., Wells, D.J.*, et al.* (2011). Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. Lancet *378*, 595-605.

Cohn, R.D., and Campbell, K.P. (2000). Molecular basis of muscular dystrophies. Muscle Nerve *23*, 1456-1471.

Collins, C.A., and Morgan, J.E. (2003). Duchenne's muscular dystrophy: animal models used to investigate pathogenesis and develop therapeutic strategies. Int J Exp Pathol *84*, 165-172.

Cooper, B.J., Winand, N.J., Stedman, H., Valentine, B.A., Hoffman, E.P., Kunkel, L.M., Scott, M.O., Fischbeck, K.H., Kornegay, J.N., Avery, R.J.*, et al.* (1988). The homologue of the Duchenne locus is defective in X-linked muscular dystrophy of dogs. Nature *334*, 154-156.

Costa, M.F., Oliveira, A.G., Feitosa-Santana, C., Zatz, M., and Ventura, D.F. (2007). Red-green color vision impairment in Duchenne muscular dystrophy. Am J Hum Genet *80*, 1064-1075.

Cotton, S.M., Voudouris, N.J., and Greenwood, K.M. (2005). Association between intellectual functioning and age in children and young adults with Duchenne muscular dystrophy: further results from a meta-analysis. Dev Med Child Neurol *47*, 257-265.

D'Souza, V.N., Nguyen, T.M., Morris, G.E., Karges, W., Pillers, D.A., and Ray, P.N. (1995). A novel dystrophin isoform is required for normal retinal electrophysiology. Hum Mol Genet *4*, 837-842.

Dangain, J., and Vrbova, G. (1984). Muscle development in mdx mutant mice. Muscle Nerve *7*, 700-704.

Duchenne (1867). The pathology of paralysis with muscular degeneration (paralysie myosclerotique), or paralysis with apparent hypertrophy. Br Med J *2*, 541-542.

Emery, A.E. (2002). The muscular dystrophies. Lancet *359*, 687-695.

England, S.B., Nicholson, L.V., Johnson, M.A., Forrest, S.M., Love, D.R., Zubrzycka-Gaarn, E.E., Bulman, D.E., Harris, J.B., and Davies, K.E. (1990). Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. Nature *343*, 180-182.

Fabb, S.A., Wells, D.J., Serpente, P., and Dickson, G. (2002). Adeno-associated virus vector gene transfer and sarcolemmal expression of a 144 kDa micro-dystrophin effectively restores the dystrophin-associated protein complex and inhibits myofibre degeneration in nude/mdx mice. Hum Mol Genet *11*, 733-741.

Fairclough, R.J., Wood, M.J., and Davies, K.E. (2013). Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches. Nat Rev Genet *14*, 373-378.

Finkel, R.S. (2010). Read-through strategies for suppression of nonsense mutations in Duchenne/ Becker muscular dystrophy: aminoglycosides and ataluren (PTC124). J Child Neurol *25*, 1158-1164.

Flanigan, K.M., Dunn, D.M., von Niederhausern, A., Soltanzadeh, P., Gappmaier, E., Howard, M.T., Sampson, J.B., Mendell, J.R., Wall, C., King, W.M.*, et al.* (2009). Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. Hum Mutat *30*, 1657-1666.

Flanigan, K.M., Dunn, D.M., von Niederhausern, A., Soltanzadeh, P., Howard, M.T., Sampson, J.B., Swoboda, K.J., Bromberg, M.B., Mendell, J.R., Taylor, L.E.*, et al.* (2011). Nonsense mutation-associated Becker muscular dystrophy: interplay between exon definition and splicing regulatory elements within the DMD gene. Hum Mutat *32*, 299-308.

Foster, H., Sharp, P.S., Athanasopoulos, T., Trollet, C., Graham, I.R., Foster, K., Wells, D.J., and Dickson, G. (2008). Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer. Mol Ther *16*, 1825-1832.

Foster, H., Popplewell, L., and Dickson, G. (2012). Genetic therapeutic approaches for Duchenne muscular dystrophy. Hum Gene Ther *23*, 676-687.

Goemans, N.M., Tulinius, M., van den Akker, J.T., Burm, B.E., Ekhart, P.F., Heuvelmans, N., Holling, T., Janson, A.A., Platenburg, G.J., Sipkens, J.A.*, et al.* (2011). Systemic administration of PRO051 in Duchenne's muscular dystrophy. N Engl J Med *364*, 1513-1522.

Gorecki, D.C., Monaco, A.P., Derry, J.M., Walker, A.P., Barnard, E.A., and Barnard, P.J. (1992). Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. Hum Mol Genet *1*, 505-510.

Gramolini, A.O., Belanger, G., and Jasmin, B.J. (2001). Distinct regions in the 3' untranslated region are responsible for targeting and stabilizing utrophin transcripts in skeletal muscle cells. J Cell Biol *154*, 1173-1183.

Greener, M.J., Sewry, C.A., Muntoni, F., and Roberts, R.G. (2002). The 3'-untranslated region of the dystrophin gene - conservation and consequences of loss. Eur J Hum Genet *10*, 413-420.

Gurvich, O.L., Tuohy, T.M., Howard, M.T., Finkel, R.S., Medne, L., Anderson, C.B., Weiss, R.B., Wilton, S.D., and Flanigan, K.M. (2008). DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy. Ann Neurol *63*, 81-89.

Hendriksen, J.G., and Vles, J.S. (2008). Neuropsychiatric disorders in males with duchenne muscular dystrophy: frequency rate of attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder, and obsessive--compulsive disorder. J Child Neurol *23*, 477-481.

Hoffman, E.P., Brown, R.H., Jr., and Kunkel, L.M. (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell *51*, 919-928.

Holder, E., Maeda, M., and Bies, R.D. (1996). Expression and regulation of the dystrophin Purkinje promoter in human skeletal muscle, heart, and brain. Hum Genet *97*, 232-239.

Howard, M.T., Anderson, C.B., Fass, U., Khatri, S., Gesteland, R.F., Atkins, J.F., and Flanigan, K.M. (2004). Readthrough of dystrophin stop codon mutations induced by aminoglycosides. Ann Neurol *55*, 422-426.

Janssen, P.M., Hiranandani, N., Mays, T.A., and Rafael-Fortney, J.A. (2005). Utrophin

deficiency worsens cardiac contractile dysfunction present in dystrophin-deficient mdx mice. Am J Physiol Heart Circ Physiol *289*, H2373-2378.

Kanamori, H., Dodson, R.E., and Shapiro, D.J. (1998). *In vitro* genetic analysis of the RNA binding site of vigilin, a multi-KH-domain protein. Mol Cell Biol *18*, 3991-4003.

Kayali, R., Ku, J.M., Khitrov, G., Jung, M.E., Prikhodko, O., and Bertoni, C. (2012). Read-through compound 13 restores dystrophin expression and improves muscle function in the mdx mouse model for Duchenne muscular dystrophy. Hum Mol Genet *21*, 4007-4020.

Khurana, T.S., and Davies, K.E. (2003). Pharmacological strategies for muscular dystrophy. Nat Rev Drug Discov *2*, 379-390.

Kinali, M., Arechavala-Gomeza, V., Feng, L., Cirak, S., Hunt, D., Adkin, C., Guglieri, M., Ashton, E., Abbs, S., Nihoyannopoulos, P.*, et al.* (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. Lancet Neurol *8*, 918-928.

Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C., and Kunkel, L.M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell *50*, 509-517.

Koo, T., Malerba, A., Athanasopoulos, T., Trollet, C., Boldrin, L., Ferry, A., Popplewell, L., Foster, H., Foster, K., and Dickson, G. (2011a). Delivery of AAV2/9-microdystrophin genes incorporating helix 1 of the coiled-coil motif in the C-terminal domain of dystrophin improves muscle pathology and restores the level of alpha1-syntrophin and alpha-dystrobrevin in skeletal muscles of mdx mice. Hum Gene Ther *22*, 1379-1388.

Koo, T., Okada, T., Athanasopoulos, T., Foster, H., Takeda, S., and Dickson, G. (2011b). Long-term functional adeno-associated virus-microdystrophin expression in the dystrophic CXMDj dog. J Gene Med *13*, 497-506.

Kornegay, J.N., Tuler, S.M., Miller, D.M., and Levesque, D.C. (1988). Muscular dystrophy in a litter of golden retriever dogs. Muscle Nerve *11*, 1056-1064.

Lederfein, D., Levy, Z., Augier, N., Mornet, D., Morris, G., Fuchs, O., Yaffe, D., and Nudel, U. (1992). A 71-kilodalton protein is a major product of the Duchenne muscular dystrophy gene in brain and other nonmuscle tissues. Proc Natl Acad Sci U S A *89*, 5346-5350.

Lemaire, C., Heilig, R., and Mandel, J.L. (1988). The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken. EMBO J *7*, 4157-4162.

Lidov, H.G., Selig, S., and Kunkel, L.M. (1995). Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. Hum Mol Genet *4*, 329-335.

Lidov, H.G., and Kunkel, L.M. (1997). Dp140: alternatively spliced isoforms in brain and kidney. Genomics *45*, 132-139.

Lu, Q.L., Rabinowitz, A., Chen, Y.C., Yokota, T., Yin, H., Alter, J., Jadoon, A., Bou-Gharios, G., and Partridge, T. (2005). Systemic delivery of antisense oligoribonucleotide restores dystrophin expression in body-wide skeletal muscles. Proc Natl Acad Sci U S A *102*, 198-203.

Malik, V., Rodino-Klapac, L.R., Viollet, L., Wall, C., King, W., Al-Dahhak, R., Lewis, S., Shilling, C.J., Kota, J., Serrano-Munuera, C*., et al.* (2010). Gentamicin-induced readthrough of stop codons in Duchenne muscular dystrophy. Ann Neurol *67*, 771-780.

Marrone, A.K., and Shcherbata, H.R. (2011). Dystrophin orchestrates the epigenetic profile of muscle cells via miRNAs. Front Genet *2*, 64.

Matoulkova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol *9*, 563-576.

Mazumder, B., Seshadri, V., and Fox, P.L. (2003). Translational control by the 3'-UTR: the ends specify the means. Trends Biochem Sci *28*, 91-98.

McCabe, E.R., Towbin, J., Chamberlain, J., Baumbach, L., Witkowski, J., van Ommen, G.J., Koenig, M., Kunkel, L.M., and Seltzer, W.K. (1989). Complementary DNA probes for the Duchenne muscular dystrophy locus demonstrate a previously undetectable deletion in a patient with dystrophic myopathy, glycerol kinase deficiency, and congenital adrenal hypoplasia. J Clin Invest *83*, 95-99.

McDonald, C.M., Henricson, E.K., Han, J.J., Abresch, R.T., Nicorici, A., Elfring, G.L., Atkinson, L., Reha, A., Hirawat, S., and Miller, L.L. (2010). The 6-minute walk test as a new outcome measure in Duchenne muscular dystrophy. Muscle Nerve *41*, 500-510.

Mendell, J.R., Rodino-Klapac, L., Sahenk, Z., Malik, V., Kaspar, B.K., Walker, C.M., and Clark, K.R. (2012). Gene therapy for muscular dystrophy: lessons learned and path forward. Neurosci Lett *527*, 90-99.

Mendell, J.R., Rodino-Klapac, L.R., Sahenk, Z., Roush, K., Bird, L., Lowes, L.P., Alfano, L., Gomez, A.M., Lewis, S., Kota, J*., et al.* (2013). Eteplirsen for the treatment of Duchenne muscular dystrophy. Ann Neurol *74*, 637-647.

Meryon, E. (1852). On granular and fatty degeneration of the voluntary muscles. Med Chir Trans *35*, 73-84 71.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome Biol *3*, REVIEWS0004.

Miura, P., and Jasmin, B.J. (2006). Utrophin upregulation for treating Duchenne or Becker muscular dystrophy: how close are we? Trends Mol Med *12*, 122-129.

Mizuno, Y., Nonaka, I., Hirai, S., and Ozawa, E. (1993). Reciprocal expression of dystrophin and utrophin in muscles of Duchenne muscular dystrophy patients, female DMD-carriers and control subjects. J Neurol Sci *119*, 43-52.

Monaco, A.P., Neve, R.L., Colletti-Feener, C., Bertelson, C.J., Kurnit, D.M., and Kunkel, L.M. (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature *323*, 646-650.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics *2*, 90-95.

Moorwood, C., Soni, N., Patel, G., Wilton, S.D., and Khurana, T.S. (2013). A cell-based high-throughput screening assay for posttranscriptional utrophin upregulation. J Biomol Screen *18*, 400-406.

Muir, L.A., and Chamberlain, J.S. (2009). Emerging strategies for cell and gene therapy of the muscular dystrophies. Expert Rev Mol Med *11*, e18.

Muntoni, F., Torelli, S., and Ferlini, A. (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol *2*, 731-740.

Muntoni, F., and Wood, M.J. (2011). Targeting RNA to treat neuromuscular disease. Nat Rev Drug Discov *10*, 621-637.

Nudel, U., Zuk, D., Einat, P., Zeelon, E., Levy, Z., Neuman, S., and Yaffe, D. (1989). Duchenne muscular dystrophy gene product is not identical in muscle and brain. Nature *337*, 76-78.

Pichavant, C., Aartsma-Rus, A., Clemens, P.R., Davies, K.E., Dickson, G., Takeda, S., Wilton, S.D., Wolff, J.A., Wooddell, C.I., Xiao, X.*, et al.* (2011). Current status of pharmaceutical and genetic therapeutic approaches to treat DMD. Mol Ther *19*, 830-840.

Pickering, B.M., and Willis, A.E. (2005). The implications of structured 5' untranslated regions on translation and disease. Semin Cell Dev Biol *16*, 39-47.

Pillers, D.A., Towbin, J.A., Chamberlain, J.S., Wu, D., Ranier, J., Powell, B.R., and McCabe, E.R. (1990). Deletion mapping of Aland Island eye disease to Xp21 between DXS67 (B24) and Duchenne muscular dystrophy. Am J Hum Genet *47*, 795-801.

Rodino-Klapac, L.R., Montgomery, C.L., Bremer, W.G., Shontz, K.M., Malik, V., Davis, N., Sprinkle, S., Campbell, K.J., Sahenk, Z., Clark, K.R.*, et al.* (2010). Persistent expression of FLAG-tagged micro dystrophin in nonhuman primates following intramuscular and vascular delivery. Mol Ther *18*, 109-117.

Rodius, F., Claudepierre, T., Rosas-Vargas, H., Cisneros, B., Montanez, C., Dreyfus, H., Mornet, D., and Rendon, A. (1997). Dystrophins in developing retina: Dp260 expression correlates with synaptic maturation. Neuroreport *8*, 2383-2387.

Schmitz, F., and Drenckhahn, D. (1997). Dystrophin in the retina. Prog Neurobiol *53*, 547-560.

Sharp, N.J., Kornegay, J.N., Van Camp, S.D., Herbstreith, M.H., Secore, S.L., Kettle, S., Hung, W.Y., Constantinou, C.D., Dykstra, M.J., Roses, A.D.*, et al.* (1992). An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. Genomics *13*, 115-121.

Sicinski, P., Geng, Y., Ryder-Cook, A.S., Barnard, E.A., Darlison, M.G., and Barnard, P.J. (1989). The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science *244*, 1578-1580.

Sonnemann, K.J., Heun-Johnson, H., Turner, A.J., Baltgalvis, K.A., Lowe, D.A., and Ervasti, J.M. (2009). Functional substitution by TAT-utrophin in dystrophin-deficient mice. PLoS Med *6*, e1000083.

Spitali, P., van den Bergen, J.C., Verhaart, I.E., Wokke, B., Janson, A.A., van den Eijnde, R., den Dunnen, J.T., Laros, J.F., Verschuuren, J.J., t Hoen, P.A.*, et al.* (2013). DMD transcript imbalance determines dystrophin levels. FASEB J *27*, 4909-4916.

Sussman, M. (2002). Duchenne muscular dystrophy. J Am Acad Orthop Surg *10*, 138-151.

Tadayoni, R., Rendon, A., Soria-Jasso, L.E., and Cisneros, B. (2012). Dystrophin Dp71: the smallest but multifunctional product of the Duchenne muscular dystrophy gene. Mol Neurobiol *45*, 43-60.

Tinsley, J.M., Fairclough, R.J., Storer, R., Wilkes, F.J., Potter, A.C., Squire, S.E., Powell, D.S., Cozzoli, A., Capogrosso, R.F., Lambert, A.*, et al.* (2011). Daily treatment with SMTC1100, a novel small molecule utrophin upregulator, dramatically reduces the dystrophic symptoms in the mdx mouse. PLoS One *6*, e19189.

Todorova, A., Todorov, T., Georgieva, B., Lukova, M., Guergueltcheva, V., Kremensky, I., and Mitev, V. (2008). MLPA analysis/complete sequencing of the DMD gene in a group of Bulgarian Duchenne/Becker muscular dystrophy patients. Neuromuscul Disord *18*, 667-670.

Valentine, B.A., Cooper, B.J., de Lahunta, A., O'Quinn, R., and Blue, J.T. (1988). Canine X-linked muscular dystrophy. An animal model of Duchenne muscular dystrophy: clinical studies. J Neurol Sci *88*, 69-81.

van Deutekom, J.C., Janson, A.A., Ginjaar, I.B., Frankhuizen, W.S., Aartsma-Rus, A., Bremmer-Bout, M., den Dunnen, J.T., Koop, K., van der Kooi, A.J., Goemans, N.M.*, et al.* (2007). Local dystrophin restoration with antisense oligonucleotide PRO051. N Engl J Med *357*, 2677-2686.

van Putten, M., Hulsker, M., Nadarajah, V.D., van Heiningen, S.H., van Huizen, E., van Iterson, M., Admiraal, P., Messemaker, T., den Dunnen, J.T., t Hoen, P.A.*, et al.* (2012). The effects of low levels of dystrophin on mouse muscle function and pathology. PLoS One *7*, e31937.

van Putten, M., Hulsker, M., Young, C., Nadarajah, V.D., Heemskerk, H., van der Weerd, L., t Hoen, P.A., van Ommen, G.J., and Aartsma-Rus, A.M. (2013). Low dystrophin levels increase survival and improve muscle pathology and function in dystrophin/utrophin double-knockout mice. FASEB J *27*, 2484-2495.

Waite, A., Tinsley, C.L., Locke, M., and Blake, D.J. (2009). The neurobiology of the dystrophin-associated glycoprotein complex. Ann Med *41*, 344-359.

Waite, A., Brown, S.C., and Blake, D.J. (2012). The dystrophin-glycoprotein complex in brain development and disease. Trends Neurosci *35*, 487-496.

Wang, B., Li, J., and Xiao, X. (2000). Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in mdx mouse model. Proc Natl Acad Sci U S A *97*, 13714-13719.

Wang, Z., Kuhr, C.S., Allen, J.M., Blankinship, M., Gregorevic, P., Chamberlain, J.S., Tapscott, S.J., and Storb, R. (2007). Sustained AAV-mediated dystrophin expression in a canine model of Duchenne muscular dystrophy with a brief course of immunosuppression. Mol Ther *15*, 1160-1166.

Wang, Z., Storb, R., Halbert, C.L., Banks, G.B., Butts, T.M., Finn, E.E., Allen, J.M., Miller, A.D., Chamberlain, J.S., and Tapscott, S.J. (2012). Successful regional delivery and long-term expression of a dystrophin gene in canine muscular dystrophy: a preclinical model for human therapies. Mol Ther *20*, 1501-1507.

Wu, J.Y., Kuban, K.C., Allred, E., Shapiro, F., and Darras, B.T. (2005). Association of Duchenne muscular dystrophy with autism spectrum disorder. J Child Neurol *20*, 790-795.

CHAPTER 1

CONSERVED REGIONS OF THE *DMD* 3' UTR REGULATE TRANSLATION AND MRNA ABUNDANCE IN CULTURED MYOTUBES

Abstract

Duchenne muscular dystrophy (DMD), a severe muscle wasting disease, is caused by mutations in the *DMD* gene, which encodes for the protein dystrophin. Its regulation is of therapeutic interest as even small changes in expression of functional dystrophin can significantly impact the severity of DMD. While tissue specific distribution and transcriptional regulation of several *DMD* mRNA isoforms have been well characterized, the posttranscriptional regulation of dystrophin synthesis is not well understood. Here, we utilize qRTPCR and a quantitative dual luciferase reporter assay to examine the effects of isoform specific *DMD* 5' UTRs and the highly conserved *DMD* 3' UTR on mRNA abundance and translational control of gene expression in C2C12 cells. The 5' UTRs were shown to initiate translation with low efficiency in both myoblasts and myotubes, whereas two large highly conserved elements in the 3' UTR, which overlap the previously described Lemaire A and D regions, increase mRNA levels and enhance translation upon differentiation of myoblasts into myotubes. The results presented here implicate an important role for *DMD* UTRs in dystrophin expression and delineate the *cis* acting elements required for the myotube-specific regulation of steady state mRNA levels

and translational enhancer activity found in the *DMD* 3' UTR.

Introduction

The DMD gene is the largest known human gene, spanning more than 2 megabases and comprising 79 exons. Tissue specific promoters drive the expression of several isoforms of dystrophin, including those expressed in brain and muscle tissue, which differ only in their first exon, and several truncated isoforms that are derived from promoters found within the *DMD* gene (Hoffman et al., 1987; Muntoni et al., 2003). The predominant muscle isoform, Dp427m, has been shown to play a structural role in maintaining muscle fiber integrity by connecting the cytoskeleton of muscle fibers to the extracellular matrix via the dystrophin associated glycoprotein complex (DGC), and to mediate signal transduction cascades through the C-terminal domain of dystrophin (Blake et al., 2002; Cohn and Campbell, 2000). Mutations in the *DMD* gene cause the dystrophinopathies, Duchenne Muscular Dystrophy (DMD) and Becker Muscular Dystrophy (BMD) (Koenig et al., 1987; Muntoni et al., 2003). Patients with DMD exhibit severe muscle degeneration, development of cardiac hypertrophy, an association with neurological disorders, and a life expectancy into the second decade of life. BMD patients exhibit muscle degeneration that occurs at a much slower rate and typically survive into late adulthood. The more severe DMD phenotype is typically caused by mutations that severely reduce or eliminate synthesis of functional dystrophin protein, such as those that disrupt the reading frame (Flanigan et al., 2009; Flanigan et al., 2011; Monaco et al., 1988; Muntoni et al., 2003), whereas the less severe BMD is caused by mutations that maintain the open reading frame and allow for residual expression of functional dystrophin protein (Flanigan et al., 2009; Flanigan et al., 2011; Monaco et al.,

1988; Muntoni et al., 2003). This observation and the finding that expression of even small amounts of dystrophin can lessen disease severity (van Putten et al., 2012; van Putten et al., 2013) has increased interest in the mechanisms that control *DMD* expression and has also led to the development of several therapeutic approaches to treat DMD patients (Fairclough et al., 2013; Muir and Chamberlain, 2009; Pichavant et al., 2011). Examples of the latter include delivery of antisense molecules designed to induce skipping of mutated exons (Cirak et al., 2011; Goemans et al., 2011; Gurvich et al., 2008; Kinali et al., 2009; Lu et al., 2005; Mendell et al., 2013; van Deutekom et al., 2007); premature stop codon suppression therapies (Barton-Davis et al., 1999; Finkel, 2010; Howard et al., 2004; Malik et al., 2010); and the delivery and expression of minidystrophin gene constructs lacking nonessential *DMD* coding exons (Fabb et al., 2002; Foster et al., 2012; Foster et al., 2008; Koo et al., 2011a; Koo et al., 2011b; Rodino-Klapac et al., 2010; Wang et al., 2000; Wang et al., 2007; Wang et al., 2012).

5' UTRs have been shown to be important for determining translation efficiency, and mutations in this region can be pathogenic (Chatterjee and Pal, 2009; Pickering and Willis, 2005). The length, GC content and secondary structures of a 5' UTR all have an impact on translation. mRNAs with high translation rates often contain 5' UTRs that are short, with low GC content, and little secondary structure, whereas mRNAs with a low translation rate, or those that are regulated in a tissue specific manner or involved in developmental processes, often contain 5' UTRs that are longer, more GC rich, and contain more secondary structure (Kochetov et al., 1998; Pickering and Willis, 2005). Most of the *DMD* 5' UTRs are longer and have a lower GC content compared to the average 5' UTR (Table 1.1) (Mignone et al., 2002; Waterston et al., 2002). Internal

Table 1.1

Properties of *DMD* 5' UTR isoforms.

Isoform	5'UTR Length	5' UTR %GC	Kozak Sequence	Isoform Expression
Average 5' UTR	210 bp	58.4%	gccrccAUGG	N/A
Dp427m	244 bp	38.9%	uucaaaAUGC	muscle
Dp427c	344 bp	44.2%	gcuggcAUGG	brain
Dp427p1	263 bp	45.2%	uuugaaAUGU	Purkinje cells
Dp427p2	703 bp	42.7%	aauguaAUGA	Purkinje cells
Dp260-1	45 bp	42.2%	auugcaAUGA	retina
Dp260-2	197 bp	34.0%	agcugaAUGA	retina
Dp140	1041 bp	43.4%	cuagaaAUGC	central nervous system, kidney
Dp116	61 bp	49.2%	cugaaaAUGU	Schwann cells
Dp71	78 bp	67.9%	gcagccAUGA	ubiquitous

Note: The 5' UTR length, percent GC content, and Kozak sequence is shown for 5' UTRs of several dystrophin isoforms along with the primary tissues expressing each isoform. The average 5' UTR length of 210 bp (Mignone et al., 2002) and percent GC content of 58.4% (Waterston et al., 2002) are shown for human 5' UTRs along with the consensus Kozak sequence (Kozak, 1987). The lengths of most *DMD* 5' UTRs are longer than average and have a GC content that is less than the average human 5' UTR. The Kozak sequence for each 5' UTR isoform and exon 1 is shown with the high impact -3 and +4 positions underlined. The *DMD* 5' UTRs have several changes compared to the consensus sequence.

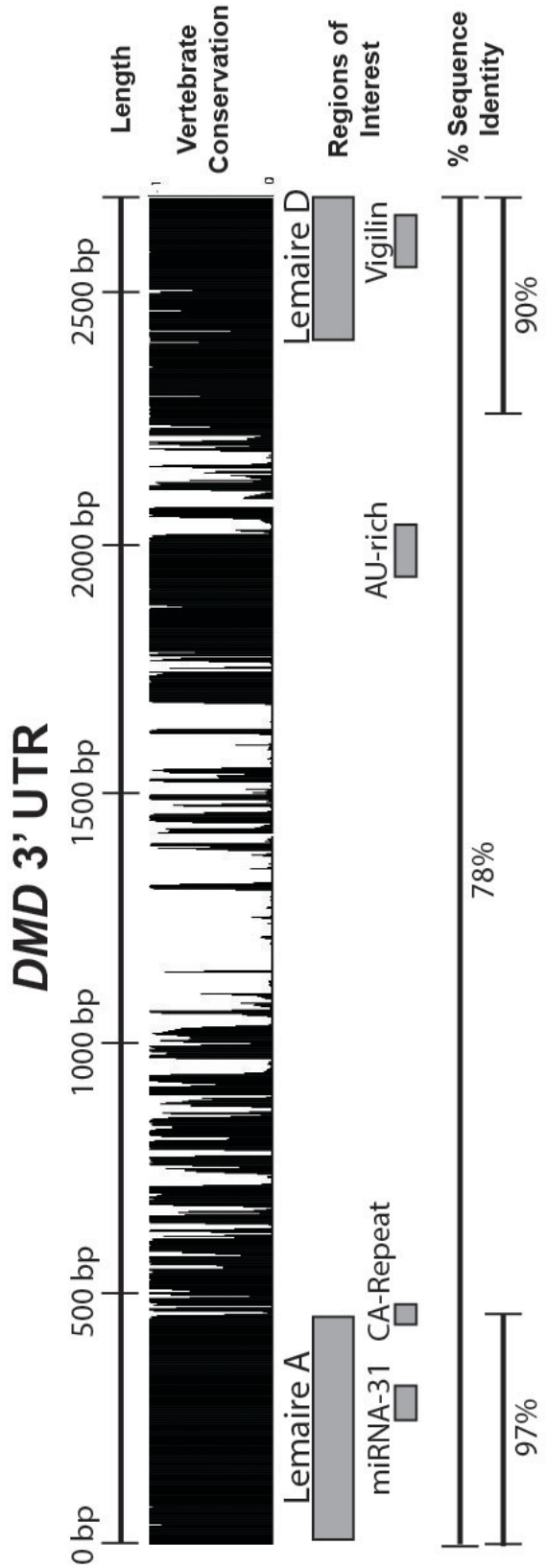
ribosome entry sites (IRES) contained within 5' UTRs have also been shown to regulate gene expression. The 5' UTR of utrophin, a homologue of dystrophin encoding a functionally related protein, was shown to regulate translation during muscle regeneration through an IRES (Miura et al., 2010; Miura et al., 2005). Likewise, the retinal dystrophin 5' UTR (Dp260) was shown to contain a cryptic intron with IRES elements that regulates translation of the R-dystrophin transcript (Kubokawa et al., 2010). The functions of the 5' UTRs for the remaining dystrophin isoforms are unknown, and may be involved in the posttranscriptional regulation of dystrophin synthesis in other tissues.

With the exception of the Dp40 isoform, all identified *DMD* mRNA isoforms encode the essential C-terminal domain and are thought to share the same 3' UTR. 3'

UTRs have been shown to regulate gene expression posttranscriptionally by altering mRNA stability, localizing mRNA, or directly affecting translation (Andreassi and Riccio, 2009; Gramolini et al., 2001; Matoulkova et al., 2012; Mazumder et al., 2003). The 2.7 kb *DMD* 3' UTR is longer than the average 3' UTR length of ~1020 nt (Mignone et al., 2002), and contains several large, highly conserved regions that may be regulating the dystrophin expression through these mechanisms (Figure 1.1). The high amount of conservation found within the *DMD* 3' UTR was first described over two decades ago (Lemaire et al., 1988). Two large, conserved regions were identified, named Lemaire A and Lemaire D, both of which are more highly conserved than the protein coding region (Figure 1.1) (Greener et al., 2002; Lemaire et al., 1988). The unusually high amount of conservation extends beyond the Lemaire sequences. For example, a sequence overlapping and extending past the Lemaire A region spans 429 nucleotides with 97.4% sequence identity between human and mouse sequences (Figure 1.1). The high conservation of these regions in the *DMD* 3' UTR is similar to the large, hyper or highly conserved elements (HCEs) identified in other 3' UTRs (Dassi et al., 2013; Siepel et al., 2005). HCEs are involved in transcriptional and posttranscriptional gene regulation (Dassi et al., 2013; Ho and Gunderson, 2011; Sathirapongsasuti et al., 2011; Siepel et al., 2005). However the mechanism of how HCEs regulate gene expression is still not understood. The extensive length and high conservation of these regions is unusual, and greater than what would be expected to retain a single binding site for RNA binding proteins or miRNAs. Nevertheless, evidence for two potential regulatory elements within these regions of the *DMD* 3' UTR has been previously reported. These include a miRNA-31 binding site within Lemaire A, thought to be involved in suppressing

Figure 1.1. Conservation of the *DMD* 3' UTR.

The conservation and regions of interest are shown for the *DMD* 3' UTR. The 2.7 kb *DMD* 3' UTR contains several large regions of high conservation. Conservation among vertebrates is shown using the PhastCons measurement of 100 vertebrate species (PhastCons track, UCSC Genome Browser). Regions of interest in the 3' UTR include the highly conserved Lemaire A and Lemaire D regions, a previously described miRNA-31 and vigilin binding sites, a variable CA repeat region, and an AU rich region. The percent sequence identity between human and mouse sequences are shown for the full 3' UTR (78%), and the regions overlapping Lemaire A (97%) or Lemaire D (90%) (See Materials and methods). The percent sequence identity in these regions are higher than the average human 3' UTR (74.7%) and the coding region (84.7%) of a gene (Waterston et al., 2002).


DMD 3' UTR
Length
Vertebrate Conservation
Regions of Interest
% Sequence Identity
0 bp
500 bp
1000 bp
1500 bp
2000 bp
2500 bp
Lemaire A
miRNA-31
CA-Repeat
AU-rich
Lemaire D
Vigilin
97%
78%
90%

dystrophin synthesis in myoblasts prior to differentiation and in dystrophic muscle (Cacchiarelli et al., 2011), and a binding site for the RNA binding protein vigilin within Lemaire D (Kanamori et al., 1998). Each of these elements spans only a small portion of the respective conserved domains and is consequently unlikely to explain the high degree of sequence conservation observed across hundreds of nucleotides in the *DMD* 3' UTR (Figure 1.1).

The results presented here demonstrate that the 5' UTRs for the full length muscle and brain isoforms initiate translation with low efficiency in C2C12 cells and that, in differentiated myotubes, the highly conserved region overlapping Lemaire A increases both mRNA levels and translation efficiency, whereas the region overlapping Lemaire D primarily affects mRNA abundance. Directed mutagenesis further demonstrates that the elements responsible for increasing dystrophin expression during differentiation span the entire Lemaire A and D domains, and that the predicted binding of vigilin and miRNA-31 does not appear to be involved in this function of the *DMD* 3' UTR in differentiated C2C12 myotubes.

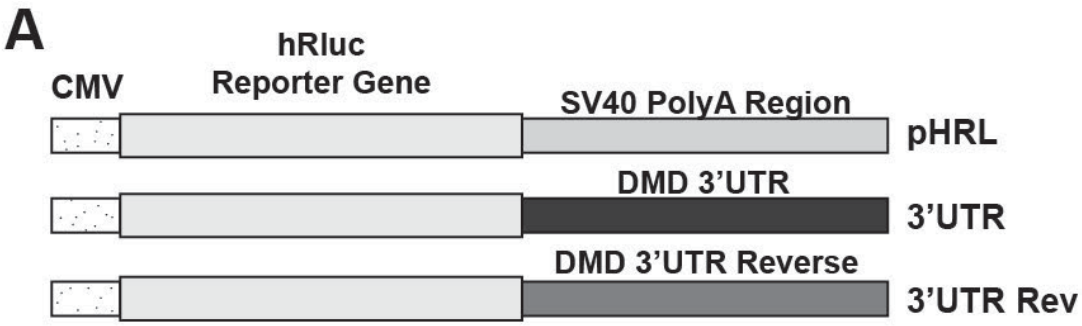
Results

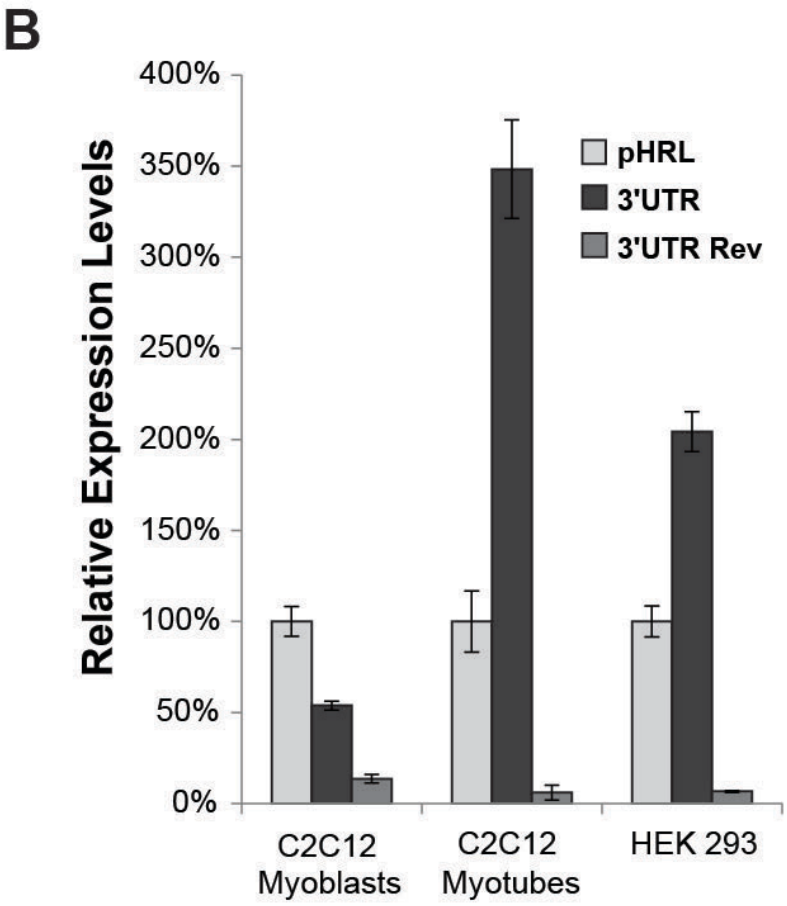
The DMD 3' UTR regulates expression during myogenesis

To investigate whether the *DMD* 3' UTR regulates gene expression, we designed a Renilla reporter construct containing the *DMD* 3' UTR downstream of the Renilla coding sequence in pHRL-CMV (3' UTR) (Figure 1.2A). This reporter construct along with controls containing either an SV40 derived 3' UTR (pHRL) or the *DMD* 3' UTR inserted in the reverse orientation (3' UTR Rev) were transfected into mouse muscle C2C12 and human embryonic kidney HEK 293 cell lines with a cotransfected pGL4.13

Figure 1.2. The *DMD* 3'UTR regulates expression in C2C12 and HEK 293 cells.

A. Diagram of the control pHRL-CMV Renilla vector (pHRL) with the SV40 PolyA region, and the pHRL-CMV vector with the *DMD* 3'UTR replacing the SV40 polyA region in either the forward (3'UTR) or reverse orientation (3'UTR Rev).
B. Relative expression levels in C2C12 and HEK 293 cells are shown. Relative expression levels are normalized to the ratio of the control Renilla pHRL vector. The average of four biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.


A
hRluc
Reporter Gene
CMV
SV40 PolyA Region
pHRL
DMD 3'UTR
3'UTR
DMD 3'UTR Reverse
3'UTR Rev


B
Relative Expression Levels
400%
350%
300%
250%
200%
150%
100%
50%
0%
pHRL
3'UTR
3'UTR Rev
C2C12
Myoblasts
C2C12
Myotubes
HEK 293

firefly luciferase vector to normalize for variation in transfection efficiency. The undifferentiated C2C12 myoblasts and HEK 293 cells were subsequently lysed and the relative amounts of Renilla and Firefly protein were measured using a Renilla/Firefly dual luciferase assay (see Materials and methods). Changes in gene expression were determined by comparing the Renilla luciferase activities expressed by each experimental construct normalized to the cotransfected Firefly luciferase control (Figure 1.2B, C2C12 Myoblasts). To determine whether the 3' UTR regulates expression during myogenesis, transfected C2C12 cells were induced to differentiate for 6 days and relative expression levels were determined as described above (Figure 1.2B, C2C12 Myotubes). The *DMD* 3' UTR increased expression by 350% and 200% compared to pHRL in C2C12 myotubes and HEK 293 cells, respectively, whereas in C2C12 myoblasts, relative expression levels decreased by 50% (Figure 1.2B). To monitor changes in expression of the pHRL-CMV (pHRL) control construct and the *DMD* 3' UTR (3' UTR) construct during differentiation, Renilla and firefly activity levels were measured every 24 hours for 7 days as the cells differentiated (Appendix C). The observed increase in expression of the 3' UTR construct (3' UTR) began 2 days after C2C12 cells were grown in differentiation media and continued to increase until the 5th day of differentiation (Appendix C). This increase in expression of the 3' UTR construct correlates with the morphology of the differentiating C2C12 cells with elongation of the myoblasts first being observed at Day 2 and large, mature myotubes being formed by Day 5 (data not shown). As a control, the DMD 3' UTR was inserted in the reverse orientation with an intact *DMD* polyadenylation signal (3' UTR Rev), and transfected into C2C12 cells or HEK 293 cells, which caused expression to decrease by more than 85% in all cell types

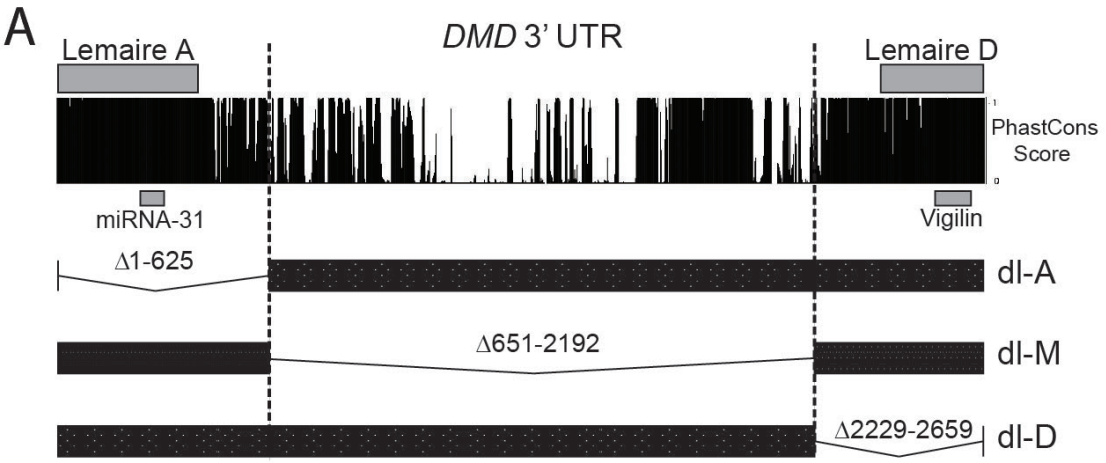
(Figure 1.2B). These experiments show that the *DMD* 3' UTR increases expression of the reporter gene in differentiating C2C12 cells and correlates with the expected increase in the muscle isoform of dystrophin in differentiated myotubes.

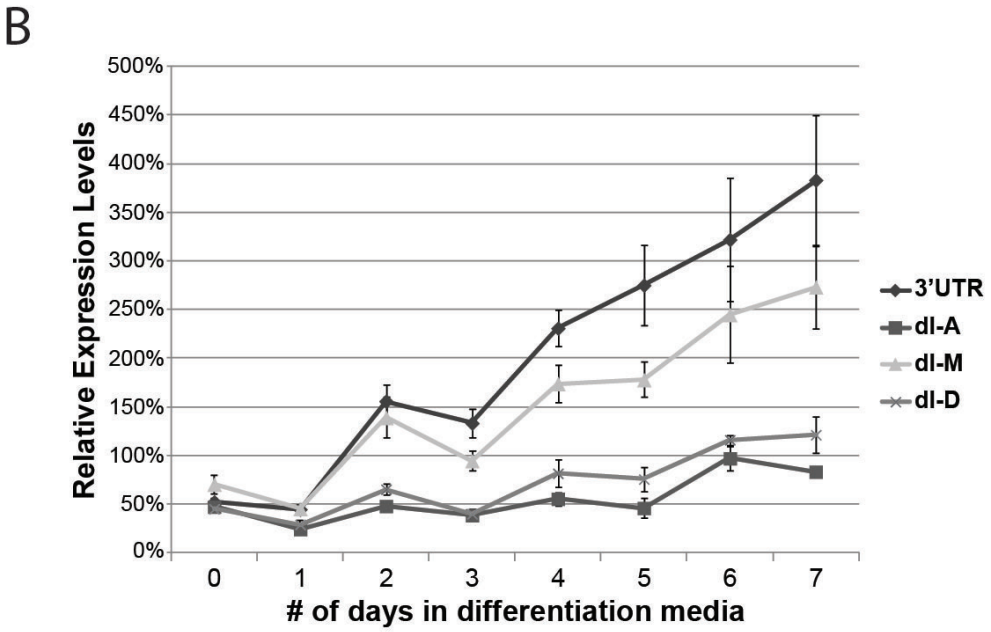
Conserved domains in the *DMD* 3' UTR regulate gene expression

Having shown that the *DMD* 3' UTR can regulate gene expression during myogenesis, we were interested in determining which regions of the DMD 3' UTR were important for this regulation. We hypothesized that the domains responsible for gene regulation are most likely contained in one of the highly conserved regions of the *DMD* 3' UTR. For this reason, we split the *DMD* 3' UTR into three segments – the first segment (A) consisting of the first 626 nucleotides of the *DMD* 3' UTR that includes Lemaire A, a CA-repeat domain, and surrounding highly conserved regions; a middle section (M) consisting of 1,542 nucleotides found in the middle of the DMD 3' UTR that contains a few regions of high conservation but consists mostly of unconserved sequences; and a third segment (D) consisting of the last 431 nucleotides of the *DMD* 3' UTR that contains Lemaire D and the surrounding highly conserved sequences (Figure 1.3A). We made three deletions in our *DMD* 3' UTR construct corresponding to each of the major sections of the *DMD* 3' UTR (dl-A, dl-M, and dl-D, Figure 1.3A) while maintaining an intact polyadenylation site and transfected these constructs into C2C12 cells. After 48 hours, the media was replaced with differentiation media and the relative expression levels for each construct were measured each day for 7 days (Figure 1.3B). As we saw before, the expression of the full length 3' UTR construct had a 6-7-fold increase in expression as the cells completed differentiation, but the constructs with the Lemaire A or Lemaire D regions deleted (dl-A, dl-D) had only about a 2-fold increase

Figure 1.3. The conserved Lemaire A and Lemaire D regions are necessary for the increase in expression during C2C12 myogenesis.

A. Diagram showing the conservation of the DMD 3'UTR and the regions deleted in each 3'UTR construct. Conservation of the *DMD* 3'UTR across vertebrates is shown using the PhastCons score across the 3'UTR (100 Vertebrate Phastcons track, UCSC genome browser). Three deletion constructs were made in the *DMD* 3'UTR that included Lemaire A (dl-A), the middle portion of the 3'UTR (dl-M), and Lemaire D (dl-D).
B. Relative expression levels were measured in C2C12 cells transfected with the full length *DMD* 3'UTR Renilla construct (3'UTR) or a deletion construct (dl-A, dl-M, or dl-D) during differentiation from myoblasts to myotubes. Expression levels were normalized to the expression of the control pHRL Renilla construct (pHRL) at each day of differentiation. The average of four biological transfection replicates is shown for each construct at each time point. Error bars equal +/- 1 standard deviation.


A
Lemaire A
DMD 3' UTR
Lemaire D
PhastCons Score
miRNA-31
Vigilin
Δ1-625
dl-A
Δ651-2192
dl-M
Δ2229-2659
dl-D


B
Relative Expression Levels
500%
450%
400%
350%
300%
250%
200%
150%
100%
50%
0%
0
1
2
3
4
5
6
7
# of days in differentiation media
3'UTR
dl-A
dl-M
dl-D

in expression. This experiment shows that the large increase in expression seen during C2C12 differentiation is dependent on the presence of both the Lemaire A and Lemaire D regions, whereas deleting the middle region of the 3' UTR (dl-M) had a smaller impact on expression (Figure 1.3B).

To determine whether the Lemaire A and Lemaire D regions were regulating expression by directly altering translation or steady state mRNA levels, we measured mRNA and protein levels for each transfection in both C2C12 myoblasts and myotubes (Figure 1.4). We measured Renilla steady state mRNA levels using qPCR relative to a cotransfected pHRL control vector, and measured Renilla protein levels for each construct using the Firefly/Renilla dual luciferase assay described above. Consistent with our previous experiment, deleting Lemaire A or Lemaire D did not have a significant effect on mRNA or protein levels in C2C12 myoblasts (Figure 1.4). Surprisingly, deleting the middle section (dl-M) resulted in a ~40-50% increase in mRNA and protein levels in myoblasts compared to the full length 3' UTR construct (3' UTR) (Figure 1.3B, Figure 1.4). In myotubes, deleting any region of the 3' UTR resulted in a significant decrease in protein expression compared to the full length 3' UTR construct, with the largest effect on expression observed when the conserved Lemaire A or Lemaire D regions were deleted. When the Lemaire A region was deleted (dl-A), steady state mRNA levels decreased by ~50% whereas luciferase activity decreased by ~75% showing that this region is critical for the increased expression seen in C2C12 myotubes by affecting both translation and steady state mRNA levels of the reporter mRNA (Figure 1.4). Deleting the Lemaire D region (dl-D) resulted in a ~50% decrease in steady state mRNA levels along with ~65% decrease in protein expression (Figure 1.4) showing that

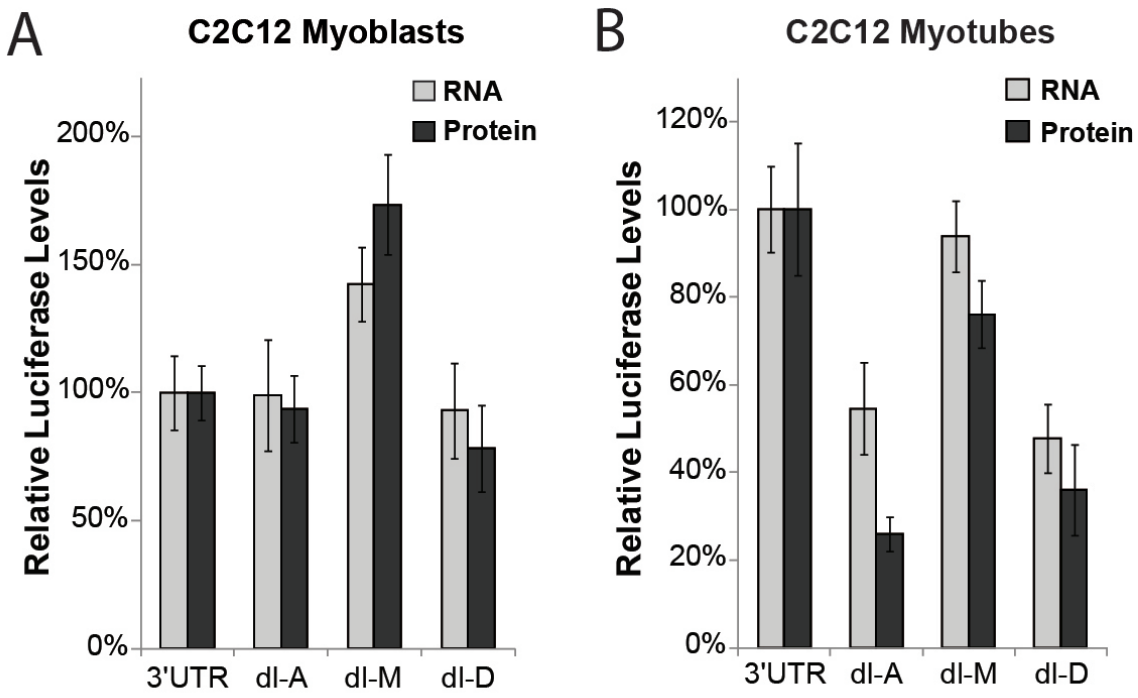


Figure 1.4. The conserved regions in the *DMD* 3 ’UTR increase translation in C2C12 myotubes.

A. Renilla mRNA and protein levels for C2C12 myoblasts transfected with the full length 3’ UTR construct (3’UTR) or deletion constructs (dl-A, dl-M, dl-D) are shown. RNA levels were measured using qPCR relative to the cotransfected pHRL control. Protein levels were measured using a Renilla/Firefly dual luciferase assay. RNA and protein levels were normalized to the amount of mRNA and protein levels of the full-length 3’UTR. The average of at least three biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.
B. Renilla mRNA and protein levels for C2C12 myotubes transfected with the full length 3’ UTR construct (3’UTR) or deletion constructs (dl-A, dl-M, dl-D) are shown. The average of at least three biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.

this region has a larger impact on maintaining mRNA steady state levels in C2C12 myotubes than on translation. To directly measure the effect these regions have on mRNA stability, we used the transcription inhibitor, Actinomycin D, to measure the half-life of our reporter constructs and found that the Lemaire A and D regions increase mRNA stability in C2C12 myotubes (Appendix A).

There is evidence that the length of a 3' UTR can affect translational efficiency and mRNA stability (Hogg and Goff, 2010; Tanguay and Gallie, 1996). To determine whether the effect of deleting regions of the *DMD* 3' UTR was due to changing the length of the 3' UTR or the absence of the region deleted, we made *DMD* 3' UTR constructs with the Lemaire A section (A) or middle section (M) of the 3' UTR reversed and transfected these constructs into C2C12 cells. Expression decreased by the same amount in C2C12 myotubes whether these regions were deleted or reversed in our constructs (Appendix C).

Lemaire A and Lemaire D regions of the *DMD* 3' UTR are necessary for increasing expression during myogenesis

It was shown previously that Lemaire A contains a predicted miRNA-31 binding site and that overexpression of miRNA-31 can inhibit dystrophin expression in cultured cells (Cacchiarelli et al., 2011). An unrelated analysis of the conserved regions in the *DMD* 3' UTR predicted the formation of a conserved 27-bp stem-loop structure within the Lemaire A region (Greener et al., 2002). To experimentally address the presence of discrete regulatory elements within the 626 nucleotide Lemaire A region, the entire region was divided into four parts of ~150 base pairs each (Aa, Ab, Ac and Ad) (Figure 1.5A). The first three regions (Aa, Ab, and Ac) overlap the previously described

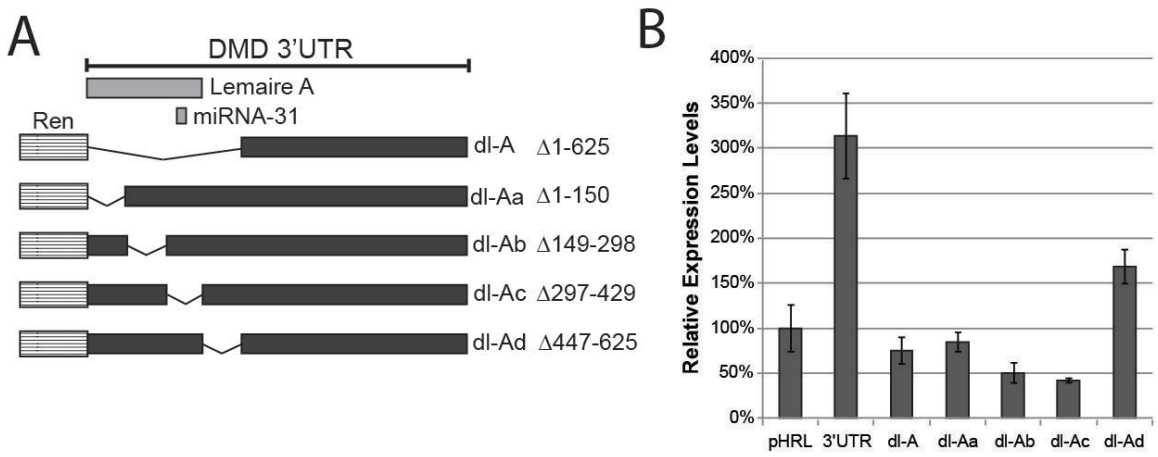


Figure 1.5. Lemaire A is necessary for high expression in C2C12 myotubes.

A. Diagram of the pHRL 3'UTR constructs with ~ 150 nucleotide portions of the Lemaire A region deleted (dl-Aa, dl-Ab, dl Ac, and dl-Ad). Lemaire A overlaps the sequences deleted in the dl-Aa, dl-Ab, and dl-Ac constructs. The miRNA-31 binding site lies within the Ac region.
B. Renilla protein expression in day six C2C12 myotubes transfected with the Lemaire A deletion constructs is shown. Relative expression levels for each construct were measured using a dual luciferase assay as previously described. The average of four biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.

Lemaire A, with the predicted miRNA-31 binding site lying within the Ac region (Figure 1.5A). The last region (Ad) contains the CA repeat and conserved sequences downstream of Lemaire A. Each of these smaller regions was deleted from the DMD 3' UTR to create the reporter constructs dl-Aa, dl-Ab, dl-Ac, dl-Ad (Figure 1.5A), and transfected into C2C12 cells. After 2 days in growth media (myoblasts) or 6 days in differentiation media (myotubes), relative expression levels were measured using the luciferase assay as previously described. In myotubes, we found that deleting any portion of Lemaire A (dl-Aa, dl-Ab, and dl-Ac) decreased expression as much as deleting the entire Lemaire A region (dl-A) (Figure 1.5B), showing that all sections of the conserved Lemaire A region must be present to increase expression in myotubes. Deleting the portion of the 3' UTR containing the CA repeat and downstream conserved sequences (dl-Ad) decreased expression by ~50% (Figure 1.5B). Consistent with our previous experiment, deleting any portion of the Lemaire A region had little impact on expression in C2C12 myoblasts compared to the full length 3' UTR construct (Figure 1.6).

To determine if the entire Lemaire D region was required for its effects on expression, we used deletion mutagenesis to make *DMD* 3' UTR constructs with ~150 bp sections of the Lemaire D region deleted (dl-Da, dl-Db, dl-Dc) (Figure 1.7A), and transfected these constructs into C2C12 cells. Deleting the entire Lemaire D region (dl-D) resulted in 65% decrease in expression compared to the full length 3' UTR construct in differentiated myotubes (Figure 1.7B). Deleting any portion of the conserved Lemaire D region (dl-Da, dl-Db, dl-Dc) resulted in a similar decrease in expression compared to deleting the entire Lemaire D region (50%, 57%, and 43% respectively) (Figure 1.7B) demonstrating that, similar to Lemaire A, the entire Lemaire D region is required for full

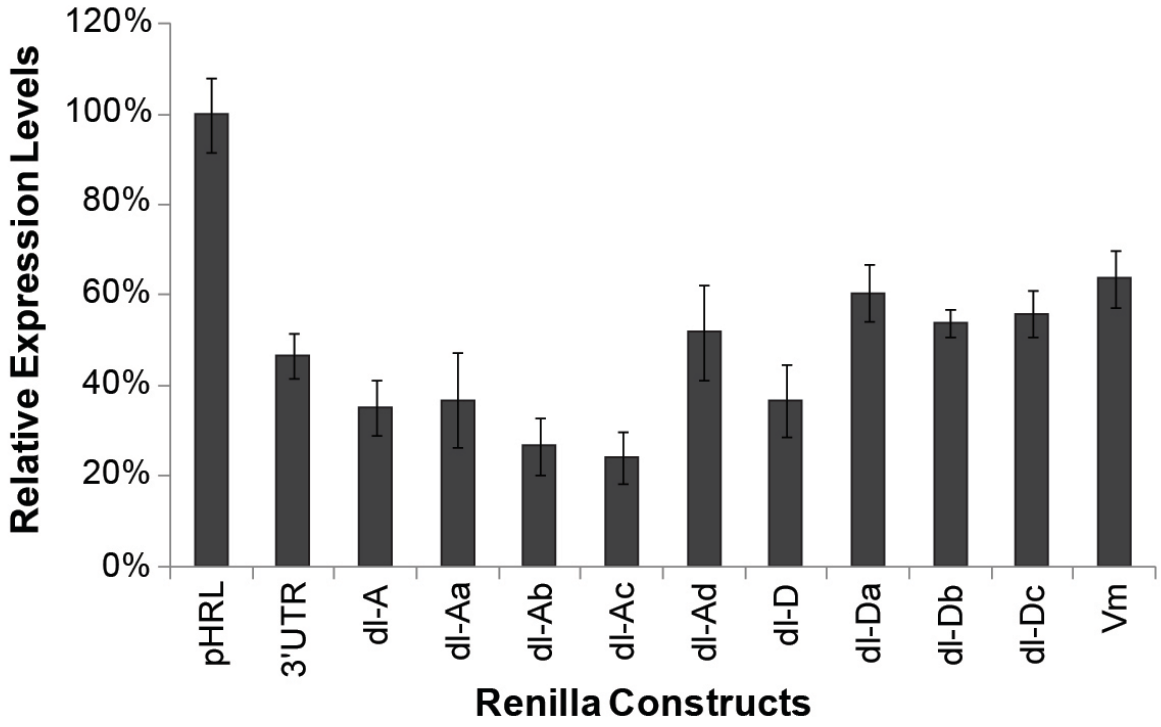


Figure 1.6. Expression of *DMD* 3' UTR constructs in C2C12 myoblasts.

Relative expression levels for the control pHRL reporter construct (pHRL), the *DMD* 3'UTR construct (3'UTR), several deletion constructs within the Lemaire A and D regions, and the vigilin mutant construct (Vm) are shown (See figure 1.5 and 1.7 for deleted regions). C2C12 myoblast cells transfected with the *DMD* 3' UTR construct (3' UTR) had a decrease in Renilla expression by 50% compared to the control pHRL reporter construct (pHRL). Deleting the Lemaire A or D regions (dl-A, dl-D) had no significant effect on expression compared to the *DMD* 3' UTR construct (3' UTR). Deleting any portion of the Lemaire A region (dl-Aa, dl-Ab, dl-Ac, and dl-Ad) or the Lemaire D region (dl-Da, dl-Db, and dl-Dc) or mutating the vigilin binding site (Vm) did not have a large impact on expression compared to the full length *DMD* 3' UTR construct (3' UTR) in C2C12 myoblasts. The average of four biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.

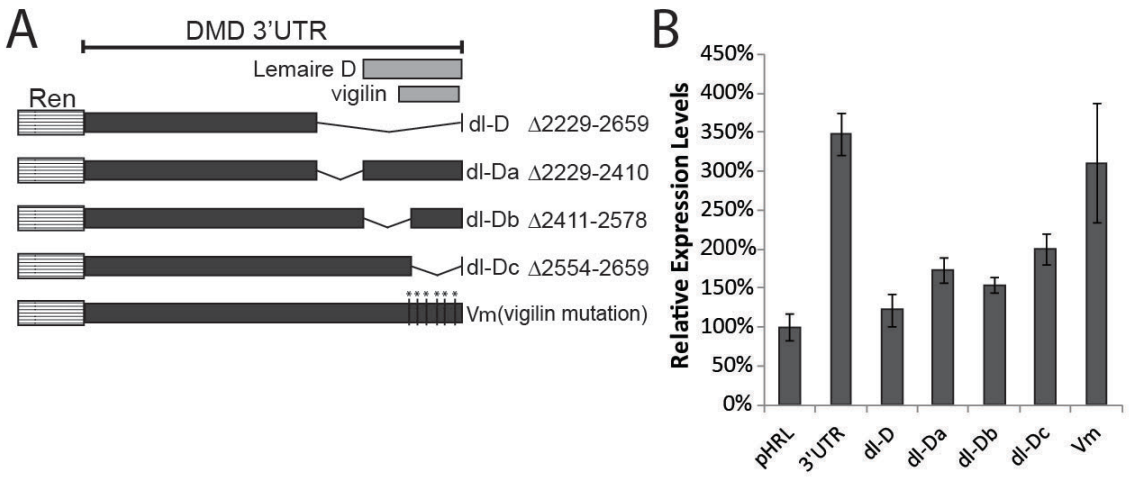


Figure 1.7. The Lemaire D region is necessary for high expression in C2C12 myotubes.

A. Diagram of the pHRL 3'UTR constructs with portions of the Lemaire D region deleted (dl-D, dl-Da, dl-Db, dl-Dc). Lemaire D and the vigilin binding site overlap the Db and Dc regions. A construct containing six point mutations in the vigilin binding site was made (Vm).
B. Relative expression levels from day 6 C2C12 myotubes transfected with the pHRL Renilla construct (pHRL), the full length 3'UTR construct (3'UTR), or Lemaire D mutant constructs (dl-Da, dl-Db, dl-Dc, and Vm containing 6 point mutations in the vigilin binding site) is shown. Relative expression levels for each construct were measured using a dual luciferase assay as previously described. The average of four biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.

expression in myotubes. Previously, it was shown that the RNA binding protein vigilin can bind *in vitro* to a predicted binding site found within Lemaire D of the *DMD* 3' UTR (overlapping the Dc region) (Figure 1.7A). This binding did not occur when six point mutations were introduced into the *DMD* 3' UTR transcript that disrupted the predicted secondary structure of the vigilin binding site (Kanamori et al., 1998). To determine whether vigilin may be regulating dystrophin expression in muscle cells, we recreated the previously described *DMD* 3' UTR mutant construct containing these six point mutations and transfected this construct into C2C12 cells. We found that mutating the vigilin binding site did not alter expression in either myoblasts (Figure 1.6) or myotubes (Figure

1.7B), suggesting that vigilin is not regulating dystrophin expression in cultured muscle cells through binding to this site.

Expression is not suppressed by miRNA-31 in C2C12 myotubes

It was shown previously that miRNA-31 could inhibit dystrophin expression by binding to a predicted miRNA-31 binding site in the *DMD* 3' UTR (Cacchiarelli et al., 2011). Undifferentiated myoblasts express miRNA-31 at high levels where this miRNA inhibits proteins involved in myogenesis and decreases in expression as the cells differentiate into myotubes (Cacchiarelli et al., 2011; Daubas et al., 2009). When we delete the Ac region overlapping the predicted miRNA-31 binding site, we see a decrease in both C2C12 myoblasts and myotubes (Figure 1.5, Figure 1.6). To determine whether miRNA-31 has an effect on our reporter construct in differentiating C2C12 cells, we made a 3' UTR construct with the 21 bp miRNA-31 binding site deleted (dl-M31), and transfected this construct into C2C12 myoblasts or myotubes (Figure 1.8). We saw no significant change in expression when the miRNA-31 binding site was deleted in our construct (dl- M31) compared to the full length 3' UTR construct (3' UTR) in C2C12 myoblasts but saw an ~40% decrease when the miRNA-31 binding site was deleted in myotubes (Figure 1.8). This result suggests that miRNA-31 is not suppressing expression in C2C12 myotubes and that there are other regulatory elements in this region increasing expression in differentiated myotubes.

An AU rich region increases expression in myotubes

The middle portion of the *DMD* 3' UTR contains only limited regions of conservation, but protein and mRNA levels significantly increased in C2C12 myoblasts

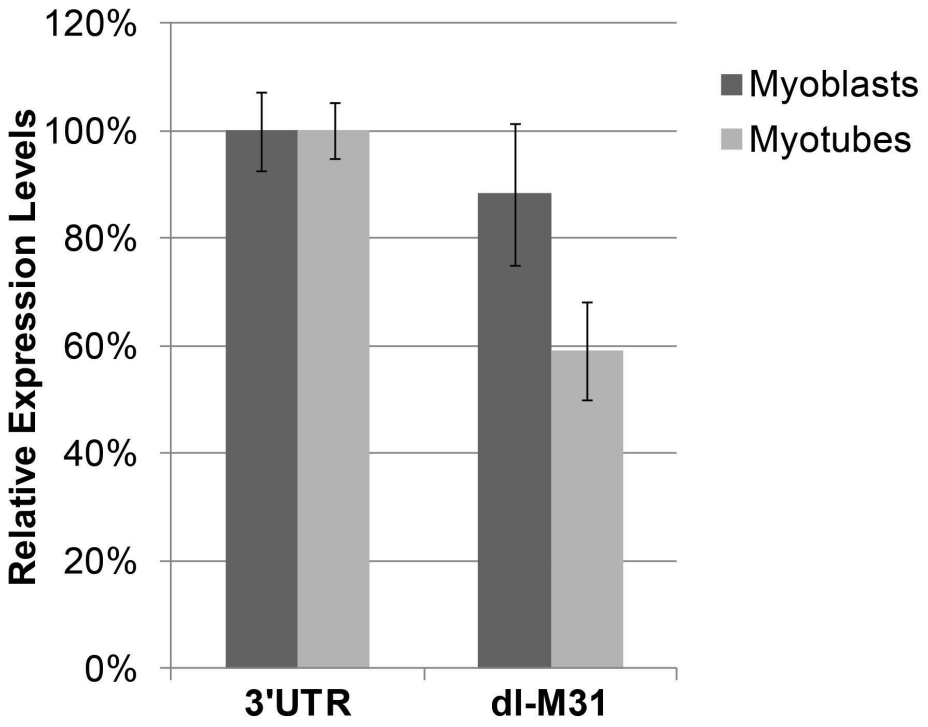


Figure 1.8. Deleting the miRNA-31 binding site decreases expression in C2C12 myotubes.

Renilla expression levels for the full length *DMD* 3' UTR construct (3'UTR) or the construct with the miRNA-31 binding site deleted (dl-M31) in transfected C2C12 myoblasts and myotubes are shown. The result of three biological replicates is shown with expression levels normalized to expression of the full length 3' UTR construct (3'UTR) in myoblasts or myotubes using a Renilla/Firefly dual luciferase assay. Error bars equal +/- 1 standard deviation.

when this region was removed, suggesting that this region may inhibit expression in myoblasts (Figure 1.4). In contrast, protein expression decreased by ~25% in myotubes when the middle region of the 3' UTR was removed (Figure 1.4). After analyzing the sequence in the middle region of the 3' UTR, we discovered a highly conserved AU rich region at the distal end of the middle section (Figure 1.1). AU rich elements in 3' UTRs have been shown to regulate mRNA stability and translation of mRNA transcripts (Barreau et al., 2005; Chakkalakal et al., 2008; Gingerich et al., 2004). To investigate

whether this sequence contained a functional AU rich element, C2C12 cells were transfected with a 3' UTR construct that contained a 262 base pair deletion spanning the conserved AU rich sequence in the middle section of the 3' UTR (dl-ARE). Deleting the AU rich sequence had no effect on protein expression of the Renilla reporter construct in myoblasts, but, in myotubes, protein expression decreased to 64% compared to the full length 3' UTR construct, similar to the protein levels seen when the entire middle region is deleted (Figure 1.9). The increase in mRNA and protein levels seen in myoblasts when the entire middle region is deleted does not appear to be due to an effect of the conserved AU rich sequence.

DMD 5' UTRs initiate translation with low efficiency

To determine the effect the Dp427m 5' UTR has on expression in muscle cells, we inserted the 5' UTR sequence from the muscle isoform of dystrophin immediately upstream of the Renilla coding sequence in our 3' UTR construct to create a Renilla construct with both the *DMD* 5' and 3' UTR (5' UTRM) (Figure 1.10A). To maintain the initiation codon sequence context for the muscle isoform of dystrophin, we also added the three amino acids of the Dp427m exon 1 transcript to the beginning of the Renilla sequence. We transfected this construct into C2C12 cells and analyzed relative expression levels in myoblasts and myotubes that were differentiated for 6 days in low serum media. We saw a significant decrease in expression when the Dp427m 5' UTR was added to the Renilla construct with 2.2% protein expression in myoblasts and 11.4% protein expression in myotubes compared to the control pHRL-CMV reporter (pHRL) (Figure 1.10B). The relative increase in expression between transfected myoblasts and myotubes is the same when the Renilla construct contains the Dp427m 5' UTR

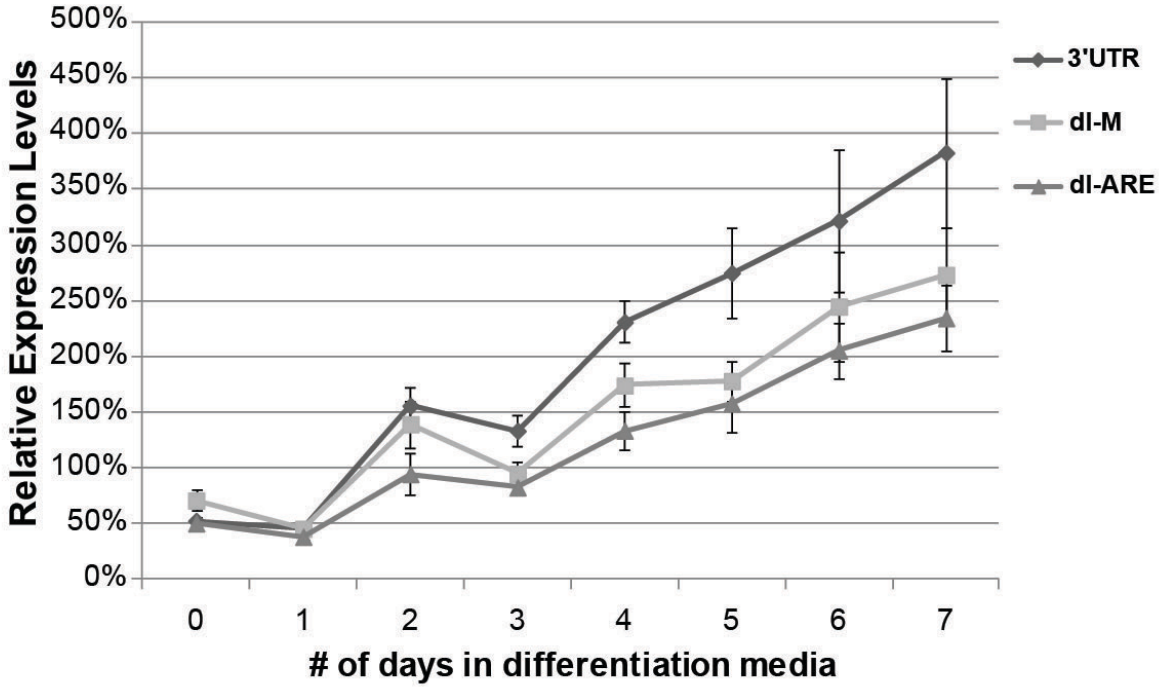
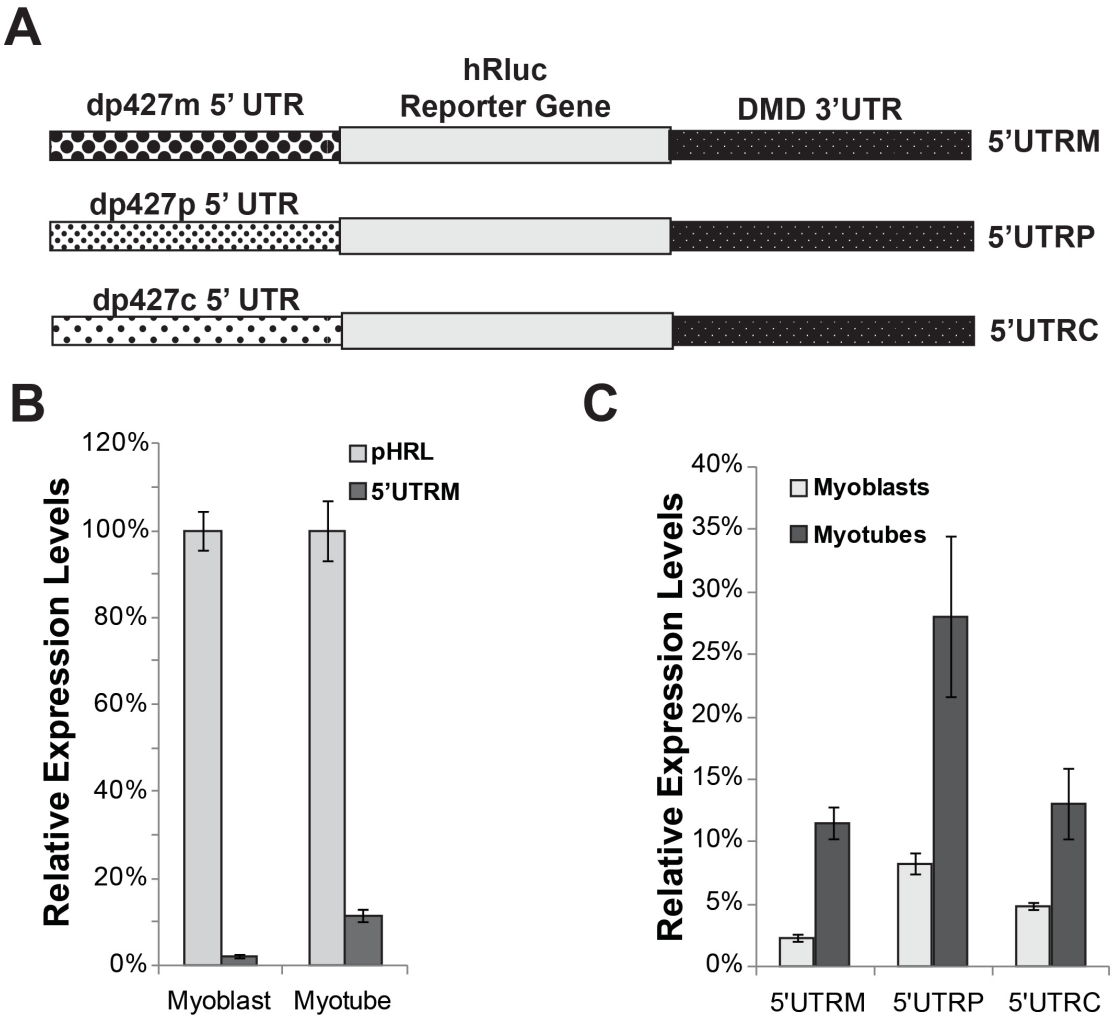


Figure 1.9. An AU rich region in the *DMD* 3'UTR decreases expression in myotubes.

Relative expression levels for C2C12 cells transfected with either the Renilla reporter construct containing the full length *DMD* 3'UTR (3'UTR), a deletion construct deleting the middle region of the *DMD* 3'UTR (dl-M), or a construct with a 262 base pair region containing an AU rich sequence found in the middle region of the 3'UTR deleted (dl-ARE) are shown for 7 days of C2C12 cell differentiation. Expression is normalized to the expression of a control Renilla construct (pHRL) at each time point. The average of four biological transfection replicates is shown for each construct at each time point. Error bars equal +/- 1 standard deviation.

Figure 1.10. The *DMD* 5'UTRs cause suboptimal translation initiation.

A. Diagram of pHRL Renilla vectors containing the *DMD* 3'UTR and the Dp427m muscle 5'UTR (5'UTRM), the Dp427p brain 5'UTR (5'UTRP), or the Dp427c brain 5'UTR (5'UTRC) inserted immediately upstream of the hRluc Renilla coding sequence in the pHRL-CMV vector.
B. Relative expression levels for C2C12 myoblasts or myotubes transfected with either the control Renilla reporter construct (pHRL) or a Renilla reporter construct containing the Dp427m 5'UTR and the full length *DMD* 3'UTR (5'UTRM) are shown.
C. Relative expression levels in C2C12 myoblasts or myotubes transfected with a reporter construct containing either the Dp427m muscle 5'UTR (5'UTRM) or the 5'UTRs of the full length Dp427p and Dp427c brain isoforms (5'UTRP, 5'UTRC) are shown. The average of four biological transfection replicates is shown for each construct at each time point. Error bars equal +/- 1 standard deviation.


A
hRluc
dp427m 5' UTR
Reporter Gene
DMD 3'UTR
5'UTRM
dp427p 5' UTR
5'UTRP
dp427c 5' UTR
5'UTRC
B
Relative Expression Levels
120%
100%
80%
60%
40%
20%
0%
pHRL
5'UTRM
Myoblast
Myotube
C
Relative Expression Levels
40%
35%
30%
25%
20%
15%
10%
5%
0%
Myoblasts
Myotubes
5'UTRM
5'UTRP
5'UTRC

(5' UTRM) compared to the 3' UTR construct lacking the 5' UTR (3' UTR). Thus the difference between myoblasts and myotube expression levels appears to be solely due to the presence of the 3' UTR. To determine if there were any differences between the different 5' UTR isoforms, we also made constructs containing the 5' UTRs from the two full length brain dystrophin isoforms (Dp427p, Dp427c) (Figure 1.10A). Similar to the results obtained for the Dp427m 5' UTR, we found a large reduction in expression in both C2C12 myoblasts and myotubes with the brain Dp427p 5' UTR (8% in myoblasts and 28% in myotubes) and Dp427c 5' UTR (5% in myoblasts and 13% in myotubes) compared to the control pHRL-CMV vector (Figure 1.10C). To determine whether the Lemaire A and Lemaire D regions had similar effects on expression in the presence of the muscle DMD 5' UTR, we used deletion mutagenesis to delete the Lemaire A and Lemaire D regions (dl-A and dl-D) in a construct containing the Dp427m 5' UTR, and transfected these constructs into C2C12 cells. The relative changes in expression when these regions were deleted was independent of whether the 5' UTR was present or absent in C2C12 myoblasts and myotubes (Figure 1.11).

Discussion

5' and 3' UTRs regulate gene expression, and mutations in these regions have been linked to a number of genetic disorders (Barrett et al., 2012; Chatterjee and Pal, 2009), including several DMD patients with mutations in the 5' and 3' UTR of the *DMD* gene. An L1 insertion in the Dp427m 5' UTR was identified in a patient with X-Linked Dilated Cardiomyopathy (XLDC) with no apparent muscular atrophy or weakness (Yoshida et al., 1998). Because no mRNA was detected in these patients, it is speculated that the insertion either disrupts transcription or stability of the muscle mRNA isoform

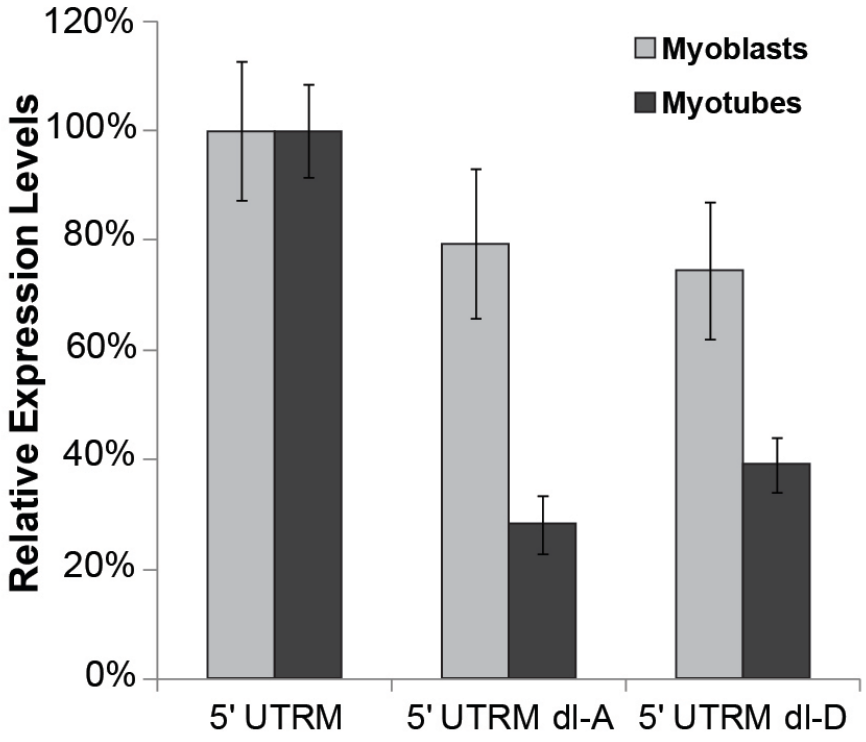


Figure 1.11. 3' UTR deletions in the Dp427m 5' UTR construct.

Relative expression levels for the Dp427m 5' UTR construct with the full length *DMD* 3' UTR (5' UTRM) or with the Lemaire A or Lemaire D regions deleted in the 3' UTR (5' UTRM dl-A and 5' UTRM dl-D, respectively) are shown for C2C12 myoblasts and myotubes (See Figure 1.3A for deletion boundaries). Deleting the Lemaire A or Lemaire D region of the 3' UTR resulted in a slight decrease in expression in C2C12 myoblasts. This is similar to the effect seen when deleting these regions when the 5' UTR is not present (Figure 1.4). In C2C12 myotubes, deleting the Lemaire A or Lemaire D regions (5' UTRM dl-A, and 5' UTRM dl-D) resulted in a significant decrease in protein expression compared to the 5' UTR construct containing the full length 3' UTR (5' UTRM). The decrease when the Lemaire A region is deleted (72%) or Lemaire D region is deleted (61%) is similar to the decrease seen in expression when the 5' UTR is not present (Figure 1.4). The effect of the highly conserved Lemaire A and D regions of the *DMD* 3' UTR on expression does not appear to be dependent on the presence of the Dp427m 5' UTR. The expression shown for each construct is relative to the amount of the 5' UTRM construct observed in myoblasts and myotubes. The average of four biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.

(Yoshida et al., 1998) and that alternate isoforms may rescue dystrophin expression and function in skeletal but not cardiac muscle. There are two examples of BMD patients with mutations in the *DMD* 3' UTR (Greener et al., 2002; McCabe et al., 1989). However, these deletions extend into the coding region of the *DMD* gene or delete other neighboring genes, and mutations contained solely within the 3' UTR have yet to be identified. Although it has been shown that altered expression levels can impact disease severity (van Putten et al., 2012; van Putten et al., 2013) and evidence presented here supports a role for the *DMD* 3' UTR in regulating expression, the identification of these mutations suggests that if pathogenic 3' UTR mutations exist, they would most likely cause the milder BMD phenotype. Another possibility is that mutations in these regions cause nonmuscular phenotypes and have been overlooked as disease causing. For example, the dystrophin glycoprotein complex is involved in brain development and function (Waite et al., 2012; Waite et al., 2009) and approximately one third of DMD patients have cognitive impairments such as reduced intelligence quotient (IQ) scores (Cotton et al., 2005), autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) (Hendriksen and Vles, 2008; Wu et al., 2005). Recently, a three base pair deletion in the coding region of the *DMD* gene was shown to cause intellectual disability, but causing no muscle phenotype (de Brouwer et al., 2014), showing that mutations which spare the muscle but have significant impacts in other tissues do exist in the *DMD* gene. Mutations in either the *DMD* 5' or 3' UTR that alter expression of one or more dystrophin isoforms could similarly have tissue specific consequences that are not currently appreciated.

Examination of the 5' UTR of the muscle (Dp427m) and brain (Dp427c and

Dp427p) isoforms of *DMD* revealed that these sequences reduced expression of a reporter construct 75-90% in both myoblasts and myotubes. Several factors may affect the efficiency of translation initiation, including the presence of upstream open reading frames or secondary structures that can limit the number of scanning ribosomes that access the initiation codon before initiation. Analysis of the *DMD* 5' UTRs reveals several upstream AUGs in the 5' UTRs of the Dp427c and Dp427p1 brain isoforms (11 and 4 AUGs, respectively) that could alter translation initiation, although there are no AUG sequences contained in the muscle Dp427m 5' UTR. Once the ribosome engages the initiation codon, positions upstream and downstream of the initiation codon determine the efficiency of productive translation initiation. The consensus translation initiation context in mammals is GCCRCCaugG, originally identified by Marilyn Kozak, and often referred to as the Kozak consensus sequence (Kozak, 1986, 1987, 1997). Deviations from this consensus, in particular at the -3 or +4 position, have been shown to reduce the efficiency of translation initiation (Kozak, 1986, 1997). The control pHRL-CMV vector matches the Kozak consensus sequence, whereas the DMD 5' UTR and exon 1 isoforms alter several key positions (Table 1.1), an observation that likely explains, at least in part, the low levels of translation initiation observed when the isoform specific DMD 5' UTRs, initiation codon, and exon 1 were placed upstream of the luciferase reporter gene.

We show that the *DMD* 3' UTR, which is common to all but one mRNA isoform, regulates expression during myogenesis. Specifically, we show that the conserved Lemaire A regulates expression by increasing mRNA abundance and translation in differentiated myotubes, and that Lemaire D and the surrounding region primarily affects mRNA abundance. Deletion mutagenesis of Lemaire A and Lemaire D shows that the

entire regions are necessary for the increase in expression observed during differentiation of myoblasts into myotubes. This result is consistent with the high level of conservation that spans these regions. The average sequence identity of 3' UTRs between human and mouse sequences is 74.7% (Waterston et al., 2002). Lemaire A has the largest effect on protein expression in C2C12 cells (the Aa, Ab, and Ac regions), and spans 429 base pairs with 97.4% sequence identity, and the Lemaire D region contains a 459 base pair sequence with 90.2% sequence identity. These regions are similar to previously described HCEs (Dassi et al., 2013; Siepel et al., 2005). HCEs are found across the genome in both coding and noncoding regions of genes (Dassi et al., 2013; Forest et al., 2007; Sathirapongsasuti et al., 2011; Siepel et al., 2005), are associated with the 3' UTRs of regulatory genes, and may be involved in posttranscriptional gene regulation, alternative splicing, and polyadenylation (Dassi et al., 2013; Ho and Gunderson, 2011; Sathirapongsasuti et al., 2011; Siepel et al., 2005). However, the mechanism of how HCEs regulate gene expression is still not understood, and the reason the extensive length and high conservation of these regions must remain intact is unknown. The existence of multiple adjacent regulatory elements and secondary structure that act cooperatively to ensure proper expression could explain the conservation of a sequence larger than a single regulatory element. One example of this is with the U1A gene that contains an approximately 53 nucleotide highly conserved region with ~91% sequence identity that binds to two U1A proteins and to the U1 snRNP splicing factor immediately upstream (Guan et al., 2007). It is speculated that the existence of multiple binding sites in this region explains the conservation seen. However, even in the presence of multiple protein binding sites, the conservation is still less than HCEs and no examples of this have been

shown for regions as large as HCEs that can span thousands of nucleotides. The essential binding of long noncoding RNAs that spans these regions is another explanation, although unlikely as long noncoding RNAs are typically less conserved than protein coding genes (Johnsson et al., 2014). Alternatively, some unknown function of the primary sequence yet to be discovered may exist.

It was shown previously that miRNA-31 could inhibit dystrophin expression by binding to a predicted miRNA-31 binding site in the *DMD* 3' UTR (Cacchiarelli et al., 2011). miRNA-31 is highly expressed in undifferentiated myoblasts where it inhibits expression of proteins involved in myogenesis, and is decreased in expression as the cells differentiate into myotubes (Cacchiarelli et al., 2011; Daubas et al., 2009). In our experiments, deleting the miRNA-31 binding site did not result in an increase in expression in C2C12 myotubes (Figure 1.8) suggesting that miRNA-31 does not repress dystrophin expression in C2C12 myotubes, and may only be repressing dystrophin expression in myoblasts and in disease tissue (Cacchiarelli et al., 2011) where miRNA-31 is highly expressed. This result does not necessarily conflict with the results of the previous study. The authors of the previous study show that miRNA-31 is overexpressed in DMD cells and that miRNA-31 can alter dystrophin expression in cells expressing high levels of miRNA-31 (Cacchiarelli et al., 2011). However, during differentiation of wild type muscle cells miRNA-31 levels decrease (Cacchiarelli et al., 2011) and our results suggest that miRNA-31 does not suppress dystrophin expression in differentiated cells where dystrophin expression is high. Instead, other regulatory elements may exist in this region that increase dystrophin expression as the cells differentiate. Our results do not negate the possible role that miRNA-31 could have on dystrophin expression in other

tissue types or in conditions where miRNA-31 is highly expressed. A conflict does exist, however, in our results in C2C12 myoblasts. The authors of the previous study show a ~30% increase in expression of their reporter construct when the miRNA-31 site was deleted (Cacchiarelli et al., 2011), whereas we saw no significant change in expression (Figure 1.8). Differences between the two experiments could account for the different results between the two studies, such as the different cell types or reporter constructs used. For example, the C2 cells used by the previous authors show diminished differentiation, lower MyoD levels, and higher protein expression levels compared to the C2C12 cells used in our experiments (Sharples et al., 2010), and the effects we measured were highly dependent on the stage of differentiation with the greatest effects seen at later stages of differentiation. The biological relevance of miRNA-31 in myoblasts is unclear because muscle dystrophin is not expressed in these cells. Although miRNA-31 could be contributing to the suppression of the muscle isoform of dystrophin in undifferentiated cells or be acting on other dystrophin isoforms.

In vitro experimental evidence has also been presented for binding of the protein vigilin to the Lemaire D region of the *DMD* 3' UTR (Kanamori et al., 1998). However, in our experiments, we saw no difference in expression when we introduced mutations shown previously to prevent vigilin binding (Kanamori et al., 1998), suggesting that vigilin is not regulating dystrophin expression in cultured muscle cells. While our experiments do not rule out the possibility that vigilin may be regulating dystrophin expression in other tissues or during other developmental stages, we find no evidence that vigilin is functioning through the 3' UTR in differentiated C2C12 myotubes.

The involvement of lncRNAs in myogenic differentiation and the potential role

they have in muscular dystrophies has become apparent in recent studies (Bovolenta et al., 2012; Cesana et al., 2011; Erriquez et al., 2013; Hube et al., 2011). Several long noncoding RNAs transcribed from within the *DMD* gene and expressed during myogenic differentiation were shown to suppress endogenous dystrophin mRNA levels in Rhabdomyosarcoma and neuroblastoma cells when overexpressed (Bovolenta et al., 2012). In addition, several of these lncRNAs suppressed the expression of a luciferase reporter driven by the Dp427m promoter in C2C12 myoblasts (Bovolenta et al., 2012). A 1,872 bp lncRNA transcribed in the antisense orientation from the middle region of the *DMD* 3' UTR was identified. While this lncRNA does not suppress dystrophin mRNA expression in these cell types (Bovolenta et al., 2012), we find that deleting the middle region of the *DMD* 3' UTR increases mRNA and protein levels by ~40-50% in undifferentiated C2C12 myoblasts, consistent with the prediction that this lncRNA could inhibit dystrophin expression in myoblasts by annealing to the complementary region of the 3' UTR. It should be noted that this effect is only observed in undifferentiated myoblasts, and that deletions within the middle region either had no effect or reduced expression in differentiated myotubes.

DMD has been a model for developing genetic therapies to treat inherited genetic mutations, including approaches using antisense oligonucleotides to cause exon skipping of dystrophin to restore the reading frame of a mutant transcript, treating patients with drugs that promote readthrough to suppress premature stop codon mutations, and expression of minidystrophin gene constructs lacking nonessential *DMD* coding exons (Fairclough et al., 2013; Muir and Chamberlain, 2009; Pichavant et al., 2011). Clinical trials using antisense oligonucleotides, or compounds that suppress premature stop codon

mutations, have shown that these compounds can restore dystrophin expression in the majority of muscle fibers of treated muscle (Goemans et al., 2011; Kinali et al., 2009; Mendell et al., 2013; van Deutekom et al., 2007). However, the total amount of dystrophin expressed in muscle tissue of these patients is highly variable between patients and typically results in less than 20% dystrophin levels compared to normal controls (Cirak et al., 2011; Goemans et al., 2011; Kinali et al., 2009; van Deutekom et al., 2007). Sequence variations in genomic regions regulating dystrophin expression, such as the *DMD* UTRs, could account for some of this variability. It has also been shown that stability of the *DMD* transcript has a large impact on dystrophin protein levels in BMD patients and increasing stability could improve the outcome of therapies such as exon skipping (Spitali et al., 2013). We have shown evidence that the *DMD* 3' UTR has an effect on mRNA stability (Appendix A) and protein levels and could have an impact on these developing therapies. In addition, restoring dystrophin expression by viral gene therapy has shown promise, but has been challenging because the 14 kb dystrophin transcript is too large to insert into viral vectors. To overcome this, minidystrophin constructs (~3-8 kb) have been designed that contain the most essential domains of the DMD coding sequence that would express functional dystrophin (Fabb et al., 2002; Foster et al., 2012; Foster et al., 2008; Koo et al., 2011a; Koo et al., 2011b; Rodino-Klapac et al., 2010; Wang et al., 2000; Wang et al., 2007; Wang et al., 2012). Our results suggest that inclusion of the regulatory elements of the *DMD* 3' UTR into these minidystrophin constructs could increase expression in differentiated muscle cells. From our experiments, we estimate that the smallest *DMD* 3' UTR that could be utilized without significantly decreasing expression would be approximately 1.2 kb and contain

the highly conserved Lemaire A and D regions of the 3' UTR.

In summary, we show the ability of the *DMD* 5' and 3' UTR to regulate gene expression during myogenesis. Patient specific variations in these regulatory elements may influence *DMD* expression and as a consequence explain some of the variation observed in disease phenotypes as well as the variable effectiveness of therapeutic approaches designed to restore endogenous dystrophin synthesis. Further, we propose that regulatory elements in the *DMD* 5' and 3' UTRs may be exploited for therapeutic purposes. For example, increasing the inherent low translation initiation efficiency imparted by the *DMD* 5' UTRs or using the regulatory elements found within the 3' UTR to increase expression in gene replacement therapy may be promising new approaches to increase dystrophin synthesis.

Materials and methods

Cell Culture

Mouse C2C12 myoblast cells were purchased from ATCC (CRL-1772). C2C12 myoblasts were cultured in growth medium consisting of Dulbecco's modified Eagle's medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Hyclone). Myogenic differentiation was induced by replacing the medium of the C2C12 myoblasts with DMEM/F-12 media supplemented with 2% horse serum (Sigma). In our studies, myoblasts were examined 2 days after being transfected, whereas myotubes were analyzed after 6 days in differentiation media unless otherwise stated. HEK 293 cells were obtained from ATCC (CRL-1573), and cultured in growth medium consisting of Minimum Essential Medium (MEM, Life Technologies) supplemented with 10% FBS (Hyclone), 1mM Sodium Pyruvate (Gibco), and 1x MEM Non Essential Amino Acids

(NEAA, Gibco). All cells were cultured at 37°C and 5% CO_2.

Generation of *DMD* 5' and 3' UTR reporter constructs and mutagenesis

The full length *DMD* 3' UTR was amplified from a European human genomic DNA sample using PCR and the following specific primers GCCGCTTCCTAGGAGGAAGTCTTTTCCACATGGC and GCTCATGGGATCCGCATGTATTACCTATTTAAAAAGTAAGTAAGTAAG that include an overhang containing the sequence of AvrII and BamH1 restriction sites, respectively. The resulting 2,811 bp PCR product was digested with XbaI and AvrII and ligated into the XbaI and BamHI restriction sites of pHRL-CMV (Promega) immediately downstream of the Renilla Luciferase ORF. The resulting plasmid contains the human *DMD* 3' UTR in place of the SV40 polyA region of the pHRL-CMV vector. The *DMD* 3' UTR was inserted in the reverse orientation as described above except that the 3' UTR from genomic DNA was amplified using a forward primer containing the BamH1 restriction site and the reverse primer containing the AvrII restriction site.

The Dp427m, Dp427c, and Dp427p1 5' UTRs and surrounding regions were amplified from a human European genomic DNA sample using PCR and primers specific to each region (Table 1.2). The resulting PCR products were used as a template for a second PCR that amplified the Dp427m, Dp427c, and Dp427p1 5 'UTRs using primers specific to each 5' UTR and containing overhang sequences with the KpnI and XhoI cut sites in the forward and reverse primers, respectively (Table 1.2). The pHRL-CMV vector containing the *DMD* 3' UTR (3' UTR construct) was amplified using primers that annealed at the beginning of the Renilla coding sequence containing an overhang with the

Table 1.2

Primers used to synthesize *DMD* 5' UTR reporter constructs.

Primers used to amplify 5' UTR genomic region from DNA		
Target	Forward Primer	Reverse Primer
dp427m 5' UTR region	CATGCTGGAAGCACAACCT	TCCTCTACTTCTTCCCACCAAA
dp427c 5' UTR region	TTTTTAGGCAGGAAGAATGC	CCAACAGCTGCTTGTTCAAA
dp427p1 5' UTR region	ATTGCTGTGAGAGGGGAGTG	CTCACAACAAAAGCCCCAAT
Primers used to amplify 5' UTR isoforms with KpnI and XhoI overhangs		
Target	Forward Primer	Reverse Primer
dp427m 5' UTR	CGTTTAGGTACCGGCATCAGTTACTGTGTTGACTCA	CGTTTACTCGAGTTTGAAAAGTGTATATCAAGGCAGC
dp427c 5' UTR	CGTTTAGGTACCCGTTAAATGCAAACGCTGC	CGTTTACTCGAGGCCAGCTGTTTTTCCTGTCA
dp427p1 5'UTR	CGTTTAGGTACCATGCTGTCTGTGAAGCTGAATCT	CGTTTACTCGAGTTCAAATTCTGCGGAGGCT
Primers used to amplify 3' UTR Renilla construct with KpnI and XhoI overhangs including isoform specific exon 1 sequences		
Target	Forward Primer	Reverse Primer
3'UTR Vector for dp427m 5' UTR	CGTTTACTCGAGATGCTTTGGATGGCTTCCAAGGTGTACGA	CGTTTAGGTACCGGTGGCTAGCCTATAGTGAGTCG
3'UTR Vector for dp427c 5' UTR	CGTTTACTCGAGATGGAAATGGCTTCCAAGGTGTACGA	CGTTTAGGTACCGGTGGCTAGCCTATAGTGAGTCG
3'UTR Vector for dp427p1 5' UTR	CGTTTACTCGAGATGTCTGAGATGGCTTCCAAGGTGTACGA	CGTTTAGGTACCGGTGGCTAGCCTATAGTGAGTCG
Primers used to remove the XhoI cut site between each 5' UTR isoform and the isoform specific exon 1		
Target	Forward Primer	Reverse Primer
dp427m 5' UTR Construct	ATGCTTTGGATGGCTTCCAA	TTTGAAAAGTGTATATCAAGGCAGC
dp427c 5' UTR Construct	ATGGAAATGGCTTCCAAGGT	GCCAGCTGTTTTTCCTGTCA
dp427p1 5' UTR Construct	ATGTCTGAGATGGCTTCCAAG	TTCAAATTCTGCGGAGGCT

KpnI cut site in the reverse primer, and an XhoI cut site and the Dp427m, Dp427c, or Dp427p exon 1 in the forward primer (Table 1.2). The amplified 5' UTR products were digested with KpnI and XhoI and ligated into the amplified 3' UTR vectors containing the exon 1 isoforms. The resulting XhoI cut site found between the *DMD* 5' UTRs and the coding sequence was removed using the Phusion Site Directed Mutagenesis Kit (Thermo Scientific), as specified by the manufacturer using phosphorylated primers specific for each construct (Table 1.2). Deletion constructs were made using the Phusion Site Directed Mutagenesis Kit (Thermo Scientific), as specified by the manufacturer using the *DMD* 3' UTR construct as a template and phosphorylated primers specific for each deletion (Table 1.3). For all subsequent experiments, plasmid DNA was prepared using QIAGEN kits. The miRNA-31 deletion construct was made using the Phusion Site Directed Mutagenesis Kit (Thermo Scientific) as specified by the manufacturer using the

Table 1.3

Phosphorylated primers used during deletion mutagenesis of the *DMD* 3' UTR.

Deletion Construct	Forward Primer	Reverse Primer
dl-A and 5' UTRM dl-A	CATTTGAGAACTGTCAGCTGAGTGG	CCTAGAATTACTGCTCGTTCTTCAGCAC
dl-M	CAAGGGGTTACTTTACATCCTACTAAGAAGTTTAAG	CCACTCAGCTGACAGTTCTCAAATG
dl-D and 5' UTRM dl-D	GTGTATCTCAATAAAGCACGCAGTTATGTTAC	CTTAAACTTCTTAGTAGGATGTAAAGTAACCCCTTG
dl-Aa	AGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAG	CCTAGAATTACTGCTCGTTCTTCAGCAC
dl-Ab	AACTACATGTAAAATCTTGATAGCTAAATAACTTGCC	CATGGTTTTTATAATATTCATACAACAAAGAGGATTAG
dl-Ac	TGTTTATAAAAACCCCTAAAAACAAAACAAAC	CTTATAACTTTTTTGTACAATTGCATAGACGTGTAAAAC
dl-Ad	CATTTGAGAACTGTCAGCTGAGTGG	AGGGGTTTTTATAAACAACTTTTTTTTATAAAGCACAC
dl-Da	GTTCTTTATAACCACCAAGTATTAAACTGT	CTTAAACTTCTTAGTAGGATGTAAAGTAACCCCTTG
dl-Db	CTTCAGTTTTACTGCATTTTCACAAC	ACAACACGAAATAATGTCCAAATTAATTAT
dl-Dc	GTGTATCTCAATAAAGCACGCAGTTATGTTAC	ACAGTGTTGGTGTTAAAACACAATATATTATG
dl-ARE	GACGCTGGACCTTTTCTTTACCC	GCCCACTCTCCAAAAGCTAATTACAC

full length *DMD* 3' UTR construct as a template and the phosphorylated primers TTTCTTTATATGGAACGCATTTTG and TTTTACATGTAGTTTTCTTATAACTTTTTTGTA. C2C12 transfections and expression level measurements using the Renilla/Firefly dual luciferase assay were done as previously described.

A *DMD* 3' UTR construct containing six point mutations in the predicted vigilin binding site as previously described (Kanamori et al., 1998) was made by amplifying the 3' UTR construct using PCR and two sets of primers containing the desired nucleotide changes. Specifically, the *DMD* 3' UTR construct was first amplified using the primers CATACTTCACCAAGTATATGCCTTACTATTATATTATAGTACTG and ACTGAAGTTTACAAAAATAATTTGTAAATGTTACAGTGTTGG (desired mutations underlined), and the purified product was amplified a second time using the phosphorylated primers AACATATCATACTTCACCAAGTATATGCCTTACTATTA and GCGAAAATGCAGTAAAACTGAAGTTTAC (desired mutations underlined). The resulting PCR product was purified and ligated using T4 DNA Ligase (New England

Biolabs) according to the manufacturer's instructions. All constructs described in the materials and methods were sequence verified.

Dual luciferase reporter assay

C2C12 myoblasts ($1x10^4$ cells) or HEK 293 cells ($4x10^4$ cells) were transfected with 25 ng of the pHRL-CMV vector (Promega) containing either the full length *DMD* 3' UTR or a mutant 3' UTR construct using 0.3uL of Lipofectamine 2000 reagent (Life Technologies) in 96 well half area plates (Corning) with 0.16 cm^2 growth area. All reactions were co transfected with 6.25 ng of the pGL4.13 [luc2/SV40] Firefly Luciferase vector (Promega) as a transfection control. To measure luciferase activity levels in myoblasts, cells were harvested 48 hours after transfection by adding 14 uL of a passive lysis buffer (Promega). To measure luciferase activity levels in myotubes, the growth media for C2C12 myoblasts was replaced 48 hours after transfection with a low serum media to induce differentiation. The differentiated myotubes were harvested after 6 days in differentiation media by adding 14 uL of a passive lysis buffer (Promega). C2C12 myoblasts and myotubes underwent one freeze-thaw cycle to aid in lysis. The bioluminescence of Renilla and firefly luciferase was measured using the Dual Luciferase Reporter Assay System (Promega) in a Veritas Microplate Luminometer (Turner Biosystems) to determine luciferase levels, and the Renilla luciferase signal was normalized to the pGL4.13 firefly luciferase signal in each well to account for variation in transfection efficiency. All transfections presented in the paper are expressed as the average +/- 1 standard deviation from a representative experiment with at least four biological replicates. For more details on transfection optimization, see Appendix C.

Determining mRNA levels of 3' UTR constructs

To determine the relative amounts of mRNA levels for the *DMD* 3' UTR construct and deletion constructs, C2C12 myoblasts ($6x10^5$ cells) were transfected with 750 ng of a 3' UTR construct and 750 ng of the control pHRL construct using 18 uL of Lipofectamine 2000 in a 24 well tissue culture plate (Falcon) with 2 cm^2 growth area. Total RNA was extracted from C2C12 cells 2 days after transfection (myoblasts) or after C2C12 cells had differentiated for 6 days in a low serum media (myotubes) using Trizol (Invitrogen) as described by the manufacturer. Each RNA sample was treated with DNAse I (Thermo Scientific) to remove any DNA contamination, and cDNA was synthesized using the SuperScript III First Strand Synthesis Kit (Life Technologies) using random hexamers. Real-time PCR was performed using the Applied Biosystems 7900HT Real time PCR system using Express SYBR GreenER qPCR SuperMix (Life Technologies) with primers optimized to amplify each construct according to the manufacturer's protocols. Results were analyzed using SDS 2.3 software and calculated using the comparative C_T ($\Delta\Delta C_T$) method by normalizing the transfected construct levels to the control pHRL cotransfected construct value at 100%. The forward and reverse primers used to amplify the cDNA of each construct are as follows: GCAACTACAACGCCTACCTT and TGGTTTGTCCAAACTCATCAA used to amplify the cotransfected pHRL vector; CTGAGGAGTTCGCTGCCTAC and GCCATGTGGAAAAGACTTCC to measure the full length 3' UTR, dl-M, and dl-D constructs; CTGAGGAGTTCGCTGCCTAC and CCCCACTCAGCTGACAGTTC to measure the dl-A construct.

Calculating percent sequence identity

To determine the percent sequence identity between human and mouse sequences in the *DMD* 3' UTR, the human and mouse Lemaire A region, Lemaire D region, or the full length 3' UTR were aligned in ClustalW. The percent sequence identity was calculated as the percent of aligned nucleotides to the total number of nucleotides in the human sequence.

Acknowledgements

I would like to thank Dr. Michael T. Howard, PhD (University of Utah, Human Genetics Department) for his continual support and guidance for this work. I thank Chris Anderson and Norma Wills for technical assistance, and Dr. Robert Weiss, PhD (University of Utah, Human Genetics Department) for comments and suggestions on this work. This work was supported in part by the National Institutes of Health NS083884 to MTH.

References

Andreassi, C., and Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. Trends Cell Biol *19*, 465-474.

Barreau, C., Paillard, L., and Osborne, H.B. (2005). AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res *33*, 7138-7150.

Barrett, L.W., Fletcher, S., and Wilton, S.D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci *69*, 3613-3634.

Barton-Davis, E.R., Cordier, L., Shoturma, D.I., Leland, S.E., and Sweeney, H.L. (1999). Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of mdx mice. J Clin Invest *104*, 375-381.

Blake, D.J., Weir, A., Newey, S.E., and Davies, K.E. (2002). Function and genetics of dystrophin and dystrophin-related proteins in muscle. Physiol Rev *82*, 291-329.

Bovolenta, M., Erriquez, D., Valli, E., Brioschi, S., Scotton, C., Neri, M., Falzarano, M.S., Gherardi, S., Fabris, M., Rimessi, P.*, et al.* (2012). The DMD locus harbours multiple long non-coding RNAs which orchestrate and control transcription of muscle dystrophin mRNA isoforms. PLoS One *7*, e45328.

Cacchiarelli, D., Incitti, T., Martone, J., Cesana, M., Cazzella, V., Santini, T., Sthandier, O., and Bozzoni, I. (2011). miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy. EMBO Rep *12*, 136-141.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell *147*, 358-369.

Chakkalakal, J.V., Miura, P., Belanger, G., Michel, R.N., and Jasmin, B.J. (2008). Modulation of utrophin A mRNA stability in fast versus slow muscles via an AU-rich element and calcineurin signaling. Nucleic Acids Res *36*, 826-838.

Chatterjee, S., and Pal, J.K. (2009). Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol Cell *101*, 251-262.

Cirak, S., Arechavala-Gomeza, V., Guglieri, M., Feng, L., Torelli, S., Anthony, K., Abbs, S., Garralda, M.E., Bourke, J., Wells, D.J.*, et al.* (2011). Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. Lancet *378*, 595-605.

Cohn, R.D., and Campbell, K.P. (2000). Molecular basis of muscular dystrophies. Muscle Nerve *23*, 1456-1471.

Cotton, S.M., Voudouris, N.J., and Greenwood, K.M. (2005). Association between intellectual functioning and age in children and young adults with Duchenne muscular dystrophy: further results from a meta-analysis. Dev Med Child Neurol *47*, 257-265.

Dassi, E., Zuccotti, P., Leo, S., Provenzani, A., Assfalg, M., D'Onofrio, M., Riva, P., and Quattrone, A. (2013). Hyper conserved elements in vertebrate mRNA 3'-UTRs reveal a translational network of RNA-binding proteins controlled by HuR. Nucleic Acids Res *41*, 3201-3216.

Daubas, P., Crist, C.G., Bajard, L., Relaix, F., Pecnard, E., Rocancourt, D., and Buckingham, M. (2009). The regulatory mechanisms that underlie inappropriate transcription of the myogenic determination gene Myf5 in the central nervous system. Dev Biol *327*, 71-82.

de Brouwer, A.P., Nabuurs, S.B., Verhaart, I.E., Oudakker, A.R., Hordijk, R., Yntema, H.G., Hordijk-Hos, J.M., Voesenek, K., de Vries, B.B., van Essen, T.*, et al.* (2014). A 3-base pair deletion, c.9711_9713del, in DMD results in intellectual disability without

muscular dystrophy. Eur J Hum Genet *22*, 480-485.

Erriquez, D., Perini, G., and Ferlini, A. (2013). Non-coding RNAs in muscle dystrophies. Int J Mol Sci *14*, 19681-19704.

Fabb, S.A., Wells, D.J., Serpente, P., and Dickson, G. (2002). Adeno-associated virus vector gene transfer and sarcolemmal expression of a 144 kDa micro-dystrophin effectively restores the dystrophin-associated protein complex and inhibits myofibre degeneration in nude/mdx mice. Hum Mol Genet *11*, 733-741.

Fairclough, R.J., Wood, M.J., and Davies, K.E. (2013). Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches. Nat Rev Genet *14*, 373-378.

Finkel, R.S. (2010). Read-through strategies for suppression of nonsense mutations in Duchenne/ Becker muscular dystrophy: aminoglycosides and ataluren (PTC124). J Child Neurol *25*, 1158-1164.

Flanigan, K.M., Dunn, D.M., von Niederhausern, A., Soltanzadeh, P., Gappmaier, E., Howard, M.T., Sampson, J.B., Mendell, J.R., Wall, C., King, W.M.*, et al.* (2009). Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. Hum Mutat *30*, 1657-1666.

Flanigan, K.M., Dunn, D.M., von Niederhausern, A., Soltanzadeh, P., Howard, M.T., Sampson, J.B., Swoboda, K.J., Bromberg, M.B., Mendell, J.R., Taylor, L.E.*, et al.* (2011). Nonsense mutation-associated Becker muscular dystrophy: interplay between exon definition and splicing regulatory elements within the DMD gene. Hum Mutat *32*, 299-308.

Forest, D., Nishikawa, R., Kobayashi, H., Parton, A., Bayne, C.J., and Barnes, D.W. (2007). RNA expression in a cartilaginous fish cell line reveals ancient 3' noncoding regions highly conserved in vertebrates. Proc Natl Acad Sci U S A *104*, 1224-1229.

Foster, H., Sharp, P.S., Athanasopoulos, T., Trollet, C., Graham, I.R., Foster, K., Wells, D.J., and Dickson, G. (2008). Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer. Mol Ther *16*, 1825-1832.

Foster, H., Popplewell, L., and Dickson, G. (2012). Genetic therapeutic approaches for Duchenne muscular dystrophy. Hum Gene Ther *23*, 676-687.

Gingerich, T.J., Feige, J.J., and LaMarre, J. (2004). AU-rich elements and the control of gene expression through regulated mRNA stability. Anim Health Res Rev *5*, 49-63.

Goemans, N.M., Tulinius, M., van den Akker, J.T., Burm, B.E., Ekhart, P.F., Heuvelmans, N., Holling, T., Janson, A.A., Platenburg, G.J., Sipkens, J.A.*, et al.* (2011). Systemic administration of PRO051 in Duchenne's muscular dystrophy. N Engl J Med

364, 1513-1522.

Gramolini, A.O., Belanger, G., and Jasmin, B.J. (2001). Distinct regions in the 3' untranslated region are responsible for targeting and stabilizing utrophin transcripts in skeletal muscle cells. J Cell Biol *154*, 1173-1183.

Greener, M.J., Sewry, C.A., Muntoni, F., and Roberts, R.G. (2002). The 3'-untranslated region of the dystrophin gene - conservation and consequences of loss. Eur J Hum Genet *10*, 413-420.

Guan, F., Caratozzolo, R.M., Goraczniak, R., Ho, E.S., and Gunderson, S.I. (2007). A bipartite U1 site represses U1A expression by synergizing with PIE to inhibit nuclear polyadenylation. RNA *13*, 2129-2140.

Gurvich, O.L., Tuohy, T.M., Howard, M.T., Finkel, R.S., Medne, L., Anderson, C.B., Weiss, R.B., Wilton, S.D., and Flanigan, K.M. (2008). DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy. Ann Neurol *63*, 81-89.

Hendriksen, J.G., and Vles, J.S. (2008). Neuropsychiatric disorders in males with duchenne muscular dystrophy: frequency rate of attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder, and obsessive--compulsive disorder. J Child Neurol *23*, 477-481.

Ho, E.S., and Gunderson, S.I. (2011). Long conserved fragments upstream of Mammalian polyadenylation sites. Genome Biol Evol *3*, 654-666.

Hoffman, E.P., Brown, R.H., Jr., and Kunkel, L.M. (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell *51*, 919-928.

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379-389.

Howard, M.T., Anderson, C.B., Fass, U., Khatri, S., Gesteland, R.F., Atkins, J.F., and Flanigan, K.M. (2004). Readthrough of dystrophin stop codon mutations induced by aminoglycosides. Ann Neurol *55*, 422-426.

Hube, F., Velasco, G., Rollin, J., Furling, D., and Francastel, C. (2011). Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. Nucleic Acids Res *39*, 513-525.

Johnsson, P., Lipovich, L., Grander, D., and Morris, K.V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta *1840*, 1063-1071.

Kanamori, H., Dodson, R.E., and Shapiro, D.J. (1998). In vitro genetic analysis of the RNA binding site of vigilin, a multi-KH-domain protein. Mol Cell Biol *18*, 3991-4003.

Kinali, M., Arechavala-Gomeza, V., Feng, L., Cirak, S., Hunt, D., Adkin, C., Guglieri, M., Ashton, E., Abbs, S., Nihoyannopoulos, P.*, et al.* (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. Lancet Neurol *8*, 918-928.

Kochetov, A.V., Ischenko, I.V., Vorobiev, D.G., Kel, A.E., Babenko, V.N., Kisselev, L.L., and Kolchanov, N.A. (1998). Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features. FEBS Lett *440*, 351-355.

Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C., and Kunkel, L.M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell *50*, 509-517.

Koo, T., Malerba, A., Athanasopoulos, T., Trollet, C., Boldrin, L., Ferry, A., Popplewell, L., Foster, H., Foster, K., and Dickson, G. (2011a). Delivery of AAV2/9-microdystrophin genes incorporating helix 1 of the coiled-coil motif in the C-terminal domain of dystrophin improves muscle pathology and restores the level of alpha1-syntrophin and alpha-dystrobrevin in skeletal muscles of mdx mice. Hum Gene Ther *22*, 1379-1388.

Koo, T., Okada, T., Athanasopoulos, T., Foster, H., Takeda, S., and Dickson, G. (2011b). Long-term functional adeno-associated virus-microdystrophin expression in the dystrophic CXMDj dog. J Gene Med *13*, 497-506.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283-292.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J Mol Biol *196*, 947-950.

Kozak, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J *16*, 2482-2492.

Kubokawa, I., Takeshima, Y., Ota, M., Enomoto, M., Okizuka, Y., Mori, T., Nishimura, N., Awano, H., Yagi, M., and Matsuo, M. (2010). Molecular characterization of the 5'-UTR of retinal dystrophin reveals a cryptic intron that regulates translational activity. Mol Vis *16*, 2590-2597.

Lemaire, C., Heilig, R., and Mandel, J.L. (1988). The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken. EMBO J *7*, 4157-4162.

Lu, Q.L., Rabinowitz, A., Chen, Y.C., Yokota, T., Yin, H., Alter, J., Jadoon, A., Bou-Gharios, G., and Partridge, T. (2005). Systemic delivery of antisense oligoribonucleotide

restores dystrophin expression in body-wide skeletal muscles. Proc Natl Acad Sci U S A *102*, 198-203.

Malik, V., Rodino-Klapac, L.R., Viollet, L., Wall, C., King, W., Al-Dahhak, R., Lewis, S., Shilling, C.J., Kota, J., Serrano-Munuera, C.*, et al.* (2010). Gentamicin-induced readthrough of stop codons in Duchenne muscular dystrophy. Ann Neurol *67*, 771-780.

Matoulkova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol *9*, 563-576.

Mazumder, B., Seshadri, V., and Fox, P.L. (2003). Translational control by the 3'-UTR: the ends specify the means. Trends Biochem Sci *28*, 91-98.

McCabe, E.R., Towbin, J., Chamberlain, J., Baumbach, L., Witkowski, J., van Ommen, G.J., Koenig, M., Kunkel, L.M., and Seltzer, W.K. (1989). Complementary DNA probes for the Duchenne muscular dystrophy locus demonstrate a previously undetectable deletion in a patient with dystrophic myopathy, glycerol kinase deficiency, and congenital adrenal hypoplasia. J Clin Invest *83*, 95-99.

Mendell, J.R., Rodino-Klapac, L.R., Sahenk, Z., Roush, K., Bird, L., Lowes, L.P., Alfano, L., Gomez, A.M., Lewis, S., Kota, J.*, et al.* (2013). Eteplirsen for the treatment of Duchenne muscular dystrophy. Ann Neurol *74*, 637-647.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome Biol *3*, REVIEWS0004.

Miura, P., Thompson, J., Chakkalakal, J.V., Holcik, M., and Jasmin, B.J. (2005). The utrophin A 5'-untranslated region confers internal ribosome entry site-mediated translational control during regeneration of skeletal muscle fibers. J Biol Chem *280*, 32997-33005.

Miura, P., Coriati, A., Belanger, G., De Repentigny, Y., Lee, J., Kothary, R., Holcik, M., and Jasmin, B.J. (2010). The utrophin A 5'-UTR drives cap-independent translation exclusively in skeletal muscles of transgenic mice and interacts with eEF1A2. Hum Mol Genet *19*, 1211-1220.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics *2*, 90-95.

Muir, L.A., and Chamberlain, J.S. (2009). Emerging strategies for cell and gene therapy of the muscular dystrophies. Expert Rev Mol Med *11*, e18.

Muntoni, F., Torelli, S., and Ferlini, A. (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol *2*, 731-740.

Pichavant, C., Aartsma-Rus, A., Clemens, P.R., Davies, K.E., Dickson, G., Takeda, S., Wilton, S.D., Wolff, J.A., Wooddell, C.I., Xiao, X.*, et al.* (2011). Current status of pharmaceutical and genetic therapeutic approaches to treat DMD. Mol Ther *19*, 830-840.

Pickering, B.M., and Willis, A.E. (2005). The implications of structured 5' untranslated regions on translation and disease. Semin Cell Dev Biol *16*, 39-47.

Rodino-Klapac, L.R., Montgomery, C.L., Bremer, W.G., Shontz, K.M., Malik, V., Davis, N., Sprinkle, S., Campbell, K.J., Sahenk, Z., Clark, K.R.*, et al.* (2010). Persistent expression of FLAG-tagged micro dystrophin in nonhuman primates following intramuscular and vascular delivery. Mol Ther *18*, 109-117.

Sathirapongsasuti, J.F., Sathira, N., Suzuki, Y., Huttenhower, C., and Sugano, S. (2011). Ultraconserved cDNA segments in the human transcriptome exhibit resistance to folding and implicate function in translation and alternative splicing. Nucleic Acids Res *39*, 1967-1979.

Sharples, A.P., Al-Shanti, N., and Stewart, C.E. (2010). C2 and C2C12 murine skeletal myoblast models of atrophic and hypertrophic potential: relevance to disease and ageing? J Cell Physiol *225*, 240-250.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S.*, et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.

Spitali, P., van den Bergen, J.C., Verhaart, I.E., Wokke, B., Janson, A.A., van den Eijnde, R., den Dunnen, J.T., Laros, J.F., Verschuuren, J.J., t Hoen, P.A.*, et al.* (2013). DMD transcript imbalance determines dystrophin levels. FASEB J *27*, 4909-4916.

Tanguay, R.L., and Gallie, D.R. (1996). Translational efficiency is regulated by the length of the 3' untranslated region. Mol Cell Biol *16*, 146-156.

van Deutekom, J.C., Janson, A.A., Ginjaar, I.B., Frankhuizen, W.S., Aartsma-Rus, A., Bremmer-Bout, M., den Dunnen, J.T., Koop, K., van der Kooi, A.J., Goemans, N.M.*, et al.* (2007). Local dystrophin restoration with antisense oligonucleotide PRO051. N Engl J Med *357*, 2677-2686.

van Putten, M., Hulsker, M., Nadarajah, V.D., van Heiningen, S.H., van Huizen, E., van Iterson, M., Admiraal, P., Messemaker, T., den Dunnen, J.T., t Hoen, P.A.*, et al.* (2012). The effects of low levels of dystrophin on mouse muscle function and pathology. PLoS One *7*, e31937.

van Putten, M., Hulsker, M., Young, C., Nadarajah, V.D., Heemskerk, H., van der Weerd, L., t Hoen, P.A., van Ommen, G.J., and Aartsma-Rus, A.M. (2013). Low dystrophin levels increase survival and improve muscle pathology and function in

dystrophin/utrophin double-knockout mice. FASEB J *27*, 2484-2495.

Waite, A., Tinsley, C.L., Locke, M., and Blake, D.J. (2009). The neurobiology of the dystrophin-associated glycoprotein complex. Ann Med *41*, 344-359.

Waite, A., Brown, S.C., and Blake, D.J. (2012). The dystrophin-glycoprotein complex in brain development and disease. Trends Neurosci *35*, 487-496.

Wang, B., Li, J., and Xiao, X. (2000). Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in mdx mouse model. Proc Natl Acad Sci U S A *97*, 13714-13719.

Wang, Z., Kuhr, C.S., Allen, J.M., Blankinship, M., Gregorevic, P., Chamberlain, J.S., Tapscott, S.J., and Storb, R. (2007). Sustained AAV-mediated dystrophin expression in a canine model of Duchenne muscular dystrophy with a brief course of immunosuppression. Mol Ther *15*, 1160-1166.

Wang, Z., Storb, R., Halbert, C.L., Banks, G.B., Butts, T.M., Finn, E.E., Allen, J.M., Miller, A.D., Chamberlain, J.S., and Tapscott, S.J. (2012). Successful regional delivery and long-term expression of a dystrophin gene in canine muscular dystrophy: a preclinical model for human therapies. Mol Ther *20*, 1501-1507.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Wu, J.Y., Kuban, K.C., Allred, E., Shapiro, F., and Darras, B.T. (2005). Association of Duchenne muscular dystrophy with autism spectrum disorder. J Child Neurol *20*, 790-795.

Yoshida, K., Nakamura, A., Yazaki, M., Ikeda, S., and Takeda, S. (1998). Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. Hum Mol Genet *7*, 1129-1132.

CHAPTER 2

ANALYSIS OF HUMAN 3' UTR VARIANTS AND THE AA REGION

Abstract

Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) are both caused by mutations in the *DMD* gene. Although a correlation between dystrophin protein or mRNA levels and disease severity is known, the posttranscriptional regulation of dystrophin has largely been unexplored. We have shown in Chapter 1 that two highly conserved elements (HCEs) overlapping Lemaire A and D in the *DMD* 3' UTR can regulate expression during differentiation in C2C12 cells. The regions of the 3' UTR affecting expression are consistent with the conservation of the 3' UTR, but the mechanism of how these regions function and how it relates to human disease are unknown. To date, no well defined human *DMD* 3' UTR mutations have been identified, although a very small percent of human DMD and BMD patients have been identified that have no mutations in the coding region of the *DMD* gene, suggesting the possibility of mutations in noncoding regions of the gene. Here, we characterize the variants in the *DMD* 3'UTR from a population of 1,222 humans that include DMD and BMD patients with no *DMD* coding mutations to determine if disease causing DMD 3' UTR mutations exist. We conclude that a disease causing mutation in the 3' UTR in this population is unlikely to exist, but that variants in the 3' UTR could modify the disease phenotype

when a second mutation is present or cause a disease phenotype in nonmuscular tissues. We also use a dual luciferase assay and an RNA protein hybridization assay to analyze the first 150 nucleotides of the HCE overlapping Lemaire A to better understand how these regions are functioning. We found that this portion of Lemaire A is sufficient in itself to increase expression in C2C12 cells although with a different expression pattern than when the entire Lemaire A is present, suggesting that Lemaire A is composed of multiple regulatory elements that work cooperatively to regulate dystrophin expression. We also present evidence that this region binds to a ~36 kDa protein and an ~85 kDa protein.

Introduction

Duchenne muscular dystrophy (DMD) and the milder Becker muscular dystrophy (BMD) are both caused by mutations in the *DMD* gene (Koenig et al., 1987; Muntoni et al., 2003). DMD patients experience progressive and severe muscle degeneration, and can die as early as the late teens or early 20s as the intercostal and cardiac muscles begin to be affected. BMD patients experience muscle degeneration and weakness similar to Duchenne muscular dystrophy but at a much slower rate, and can typically survive into the fourth or fifth decades of life. The severity of disease amongst BMD patients is highly variable. Onset of clinical symptoms can begin around 12 years of age, but some patients do not show symptoms until late adulthood. Likewise, human clinical trials using compounds to restore dystrophin expression in muscle tissue have shown success, but the restored dystrophin levels are highly variable between patients and typically result in less than 20% dystrophin levels compared to normal controls (Cirak et al., 2011; Goemans et al., 2011; Kinali et al., 2009; van Deutekom et al., 2007). There is also a

correlation between dystrophin protein and mRNA levels and disease severity (Spitali et al., 2013; van Putten et al., 2012; van Putten et al., 2013). These observations have important implications, not only for determining the prognosis of a patient, but also in the effectiveness of developing therapeutic treatments aimed at restoring dystrophin expression in DMD patients (Fairclough et al., 2013; Muir and Chamberlain, 2009; Pichavant et al., 2011). Variation in regulatory factors involved in dystrophin expression could account for some of the variation in disease severity seen in human patients and impact the effectiveness of developing treatments.

While tissue specific distribution and transcriptional regulation of *DMD* mRNA isoforms have been well characterized, the posttranscriptional regulation of dystrophin synthesis is not well understood. 3' UTRs are known to regulate gene expression posttranscriptionally by altering mRNA stability, localizing mRNA, or directly affecting translation (Andreassi and Riccio, 2009; Gramolini et al., 2001; Matoulkova et al., 2012; Mazumder et al., 2003), and we were able to show that the *DMD* 3'UTR regulates expression during C2C12 cell differentiation (Chapter 1). We identified two highly conserved regions that overlap and extend beyond the previously described Lemaire A and D regions of the 3' UTR that increase translation and mRNA abundance in C2C12 myotubes (Chapter 1). Using deletion mutagenesis, we were unable to narrow down the regulatory sequences within these regions below ~400 base pairs because deleting any portion of the highly conserved Lemaire A or D regions resulted in decreased expression equal to or greater than the expression when the entire element was deleted. Thus, the entire element must remain intact for proper function. It is unclear from these experiments whether the entire sequence encompasses a single regulatory element that

must remain intact to increase expression or if these regions contain several regulatory elements acting cooperatively to regulate dystrophin expression.

Regions of similar size and conservation as these highly conserved regions in the *DMD* 3' UTR have been described in the literature, although the nomenclature of these regions has not been well defined. These regions have been referred to as highly conserved elements (HCEs) (Siepel et al., 2005); hyper conserved elements (HCEs) (Dassi et al., 2013); conserved fragments (CFs) (Ho and Gunderson, 2011); and ultra conserved elements (UCEs) (Bejerano et al., 2004), but, in general, refer to sequences that are 100% or nearly 100% conserved between human and mouse sequences and span anywhere from 50 nucleotides to several thousand nucleotides. Throughout this dissertation, I will refer to these elements as highly conserved elements (HCEs) defined as unusually large sequences in the genome that exceed the conservation seen in the protein coding regions of genes, which encompasses the range of definitions that have been used to describe these elements.

A genome wide analysis that ranked HCEs based on conservation found that of the top 5000 elements, 42% overlapped protein coding exons, 9% overlapped 5' UTRs, 16% overlapped 3' UTRs, 19% were contained within introns, and 32% were in unannotated regions (Siepel et al., 2005). This analysis identified 10 HCEs in the *DMD* 3' UTR including two large HCEs that span the Lemaire A and Lemaire D regions (Siepel et al., 2005). This study also showed that 3' UTRs throughout the genome were enriched for HCEs. 3' UTRs accounted for 5.6% of all conserved bases analyzed in this study, but 9.6% of bases in the top 5000 conserved elements, and 14.3% of bases in the top 100 elements were found in 3' UTRs (Siepel et al., 2005). Despite the enrichment of

HCEs in 3' UTRs, they are still rarely seen. Only 1.8% of all human 3' UTRs contain at least one HCE, and collectively, HCEs account for 0.47% of total 3' UTR sequence (Dassi et al., 2013). 3' UTR HCEs have an average length of 100 bases, with 77% of all HCEs being shorter than 100 bases, and only 4.5% of them being larger than 500 bases (Dassi et al., 2013). 3' UTR HCEs tend to localize into clusters when multiple HCEs are present in the same 3' UTR, and are more enriched at the beginning of the 3' UTR with 25% of HCEs starting within the first 10% of the 3' UTR (Dassi et al., 2013). The HCE overlapping Lemaire A in the *DMD* 3' UTR spans 429 base pairs with 97.4% sequence identity between human and mouse sequences with sections up to 124 base pairs with 100% conservation, and begins immediately after the stop codon. The HCE overlapping Lemaire D at the end of the 3' UTR is less conserved with 90% sequence identity spanning a 459 base pair region with segments up to 49 base pairs with 100% conservation, although the sequence identity increases to 97.8% for the last 184 base pairs of this region.

The mechanism of how 3' UTR HCEs function is not well understood, but HCEs have been shown to be enriched in genes involved in the ubiquitin cycle, RNA binding, mRNA metabolism, and mRNA processing (Siepel et al., 2005), and functional evidence suggests that HCEs are involved in posttranscriptional gene regulation, alternative splicing, and polyadenlyation (Dassi et al., 2013; Ho and Gunderson, 2011; Sathirapongsasuti et al., 2011; Siepel et al., 2005). It is not clear why such large noncoding regions remain conserved, but analysis of these regions suggests that 3' UTR HCEs contain far more secondary structure than random 3' UTR sequences (Dassi et al., 2013; Siepel et al., 2005), and may contain biologically relevant structure. These regions

may also function by interacting with miRNA or RNA binding proteins. There is evidence that miRNAs regulate HCEs and genes containing HCEs (Calin et al., 2007; Sathirapongsasuti et al., 2011). However, 3' UTR sequences lying outside of HCEs are more enriched in predicted miRNA binding sites than 3' UTR HCEs with only 1.6% of HCEs analyzed in one large study containing at least one predicted miRNA binding site (Dassi et al., 2013). There is evidence that 3' UTR HCEs bind to proteins. There is enrichment for experimentally identified binding sites for RNA binding proteins in 3' UTR HCEs, and at least one example where the RNA binding protein HuR regulates a network of genes by binding to 3' UTR HCEs (Dassi et al., 2013). The only experimental evidence of a miRNA or RNA binding protein binding to the *DMD* 3' UTR HCEs is for miRNA-31 and RNA binding protein vigilin, but our results in Chapter 1 suggest that these factors are not involved in the increased expression seen during C2C12 differentiation. However, there are several predicted RNA binding protein and miRNA binding sites in the conserved regions of the *DMD* 3' UTR (Appendix B) that could potentially be functional in regulating dystrophin expression.

Mutations in 3' UTRs have been associated with several diseases in humans, including point mutations and small insertions/deletions that are disease causing, disease predisposing, or disease modifying (Chen et al., 2006a, b). 3' UTR mutations have been shown to cause disease by altering the polyadenylation signal, predicted secondary structure, or predicted binding sites for miRNAs or RNA binding proteins (Chen et al., 2006a, b). Select examples in humans where even a small change in a 3' UTR can cause disease include a T>C point mutation in the selenocysteine insertion sequence (SECIS) in the 3' UTR of the SEPN1 gene that causes a mild form of rigid spine muscular dystrophy

(Allamand et al., 2006), a C>T polymorphism in an AU rich element in the TGFB3 gene shown to cause atypical arrhythmogenic right ventricular dysplasia (ARVD) and to increase reporter gene activity in C2C12 myoblasts (Beffagna et al., 2005), and, finally, a G>A transition in a predicted binding site for miRNA-189 in the 3' UTR of the SLITRK1 gene that was associated with Tourette's Syndrome and shown to decrease expression of a reporter construct (Abelson et al., 2005). For a comprehensive list of disease associated 3' UTR mutations, see the review (Chen et al., 2006a, b).

The effect the *DMD* 3' UTR has on human disease is unknown. Thus far, no well defined disease causing mutations have been identified in the *DMD* 3' UTR in human DMD or BMD patients. There are two examples of BMD patients (Greener et al., 2002; McCabe et al., 1989) and two examples of DMD patients (Pillers et al., 1990; Todorova et al., 2008) with deletions that span the *DMD* 3' UTR. However, these deletions extend into the coding region of the *DMD* gene, delete other neighboring genes, or have not been fully analyzed to determine the extent of the deletion, and disease causing mutations shown to be contained solely within the 3' UTR have yet to be identified. However, there is a small group of human patients (0.2% in one large study) (Flanigan et al., 2009) where dystrophin expression is absent, but no *DMD* mutations exist in the coding region of the gene. It is possible that a subset of these patients has disease causing mutations in noncoding regions of the gene that has been overlooked by standard mutation screens.

Here, we analyzed *DMD* 3' UTR sequences from a population of 1,222 humans that include several DMD and BMD patients with no *DMD* coding mutations, and characterize the variants found in this population and other publicly available databases to determine the likelihood of a disease causing 3' UTR mutation existing in this group.

To better understand how the Lemaire A region regulates expression, we also conduct a more thorough analysis of the first 150 base pairs of the Lemaire A region (termed the Aa region). We have shown in Chapter 1 that this Aa region is necessary for the increased expression of a reporter construct in C2C12 myotubes, and that the decrease in expression when this region is deleted is equal to the decrease when the entire Lemaire A region is deleted. We investigate whether the Aa region alone is sufficient to regulate expression in C2C12 cells. We also attempt to narrow down the regulatory sequences contained in this region by subdividing the region into smaller portions, and investigate the potential of RNA binding proteins to bind to this region.

We show that the first 150 base pairs of the Lemaire A region are sufficient to increase expression in C2C12 cells, and show evidence that a ~36 kDa protein and an ~85 kDa protein can bind to the distal portion of the Aa region, suggesting that the Lemaire A region contains multiple regulatory elements and may be interacting with RNA binding proteins. In our analysis of human 3' UTR variants, we conclude that a disease causing mutation in the *DMD* 3' UTR is unlikely to exist in the group of patients analyzed. We show that the only human variant in this population found at a predicted protein binding site in the Aa region in our DMD population (where a T nucleotide is inserted at a predicted HuR binding site) does not alter expression in C2C12 cells. However, we could not rule out the possibility that variants contained within the highly conserved regions of the 3' UTR could act as modifying mutations that alter the disease phenotype when other mutations are present or cause a disease phenotype in nonmuscular tissues.

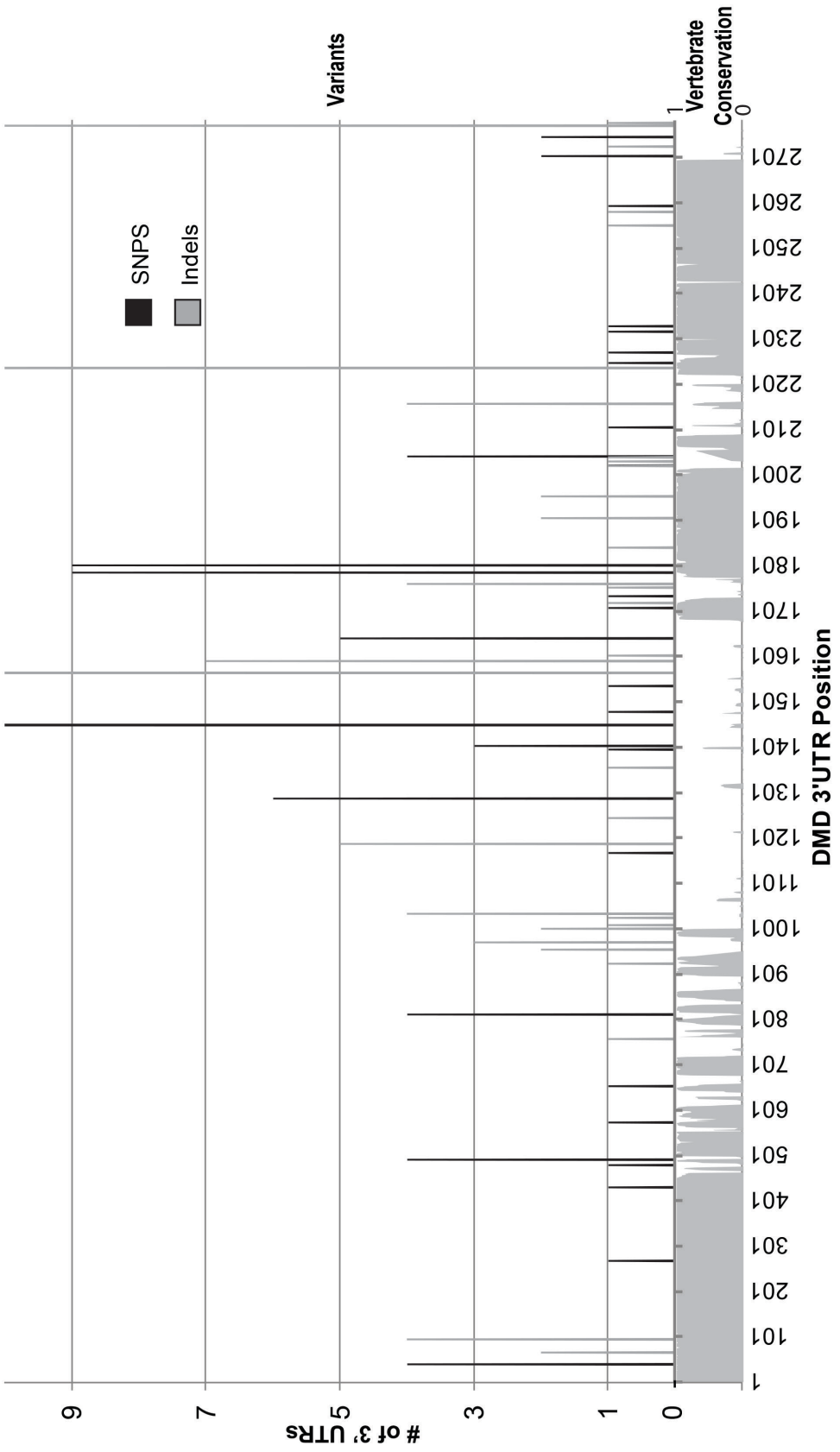
Results

Characterization of *DMD* 3' UTR variants in humans

To characterize the human variants in the *DMD* 3' UTR and to determine if disease causing mutations exist in the *DMD* 3' UTR, we obtained sequences from Dr. Robert Weiss (University of Utah, Human Genetics Department) of 1,222 human *DMD* 3' UTRs from three groups of patients that included DMD and BMD patients that had no *DMD* coding mutations (NODCM). Sequences from 410 patients were obtained from patients enrolled in the United Dystrophinopathy Project (UDP), a registry of DMD and BMD patient information with identified *DMD* mutations that is managed by Dr. Kevin Flanigan (Nationwide Children's Hospital, Columbus, OH). This group contained 51 NODCM patients. Sequences from 569 patients were obtained from clinical referrals from doctors who had requested *DMD* mutation screening for individuals suspected to be dystrophic, and this group contained 189 NODCM patients. Sequences from 243 patients were obtained from a newborn screening study that screened for *DMD* mutations in infants with high creatine kinase (CK) levels (Mendell et al., 2012). Most of these infants were not dystrophic and had no *DMD* mutations (Mendell et al., 2012) with 228 of these patients identified as NODCM patients. The sequences from all patients were analyzed and any discrepancy between a patient sequence and a reference sequence (NM_004006.2) was visually inspected to determine the quality of the trace. All variants identified were named using the Human Genome Variation Society (HGVS) guidelines and recommendations for variant nomenclature (den Dunnen and Antonarakis, 2000). We verified 1,377 discrepancies that consisted of 84 unique variants in the *DMD* 3' UTR across 435 patients (Figure 2.1, Appendix B). We also searched for *DMD* 3' UTR

Figure 2.1. Human variation in the *DMD* 3' UTR.

Diagram of the *DMD* 3' UTR SNPs and small insertions and deletions (Indels) found in 444 human DMD patients is shown. The cDNA *DMD* 3' UTR position along with the PhastCons conservation scores along the length of the 3' UTR is depicted on the X axis. Peak heights are equal to the number of 3' UTRs found for each variant. Variants found in more than 10 3' UTRs are capped at a peak height of 10.


Variants
Vertebrate Conservation
SNPS
Indels
# of 3' UTRs
DMD 3'UTR Position

variants reported in publicly available databases, and found a total of 83 variants in the current version of dbSNP (Sherry et al., 2001), 40 variants in the 1000 Genome database (Abecasis et al., 2012), and 38 variants reported in the Leiden Database (Aartsma-Rus et al., 2006) (Appendix B). There were 56 additional variants found in these public databases that were not found in our patient population, and 40 variants found exclusively in the patients we analyzed. In total, 140 unique 3' UTR variants were identified in our patient population and all databases searched (Appendix B). Although most of these variants are found in unconserved regions of the *DMD* 3' UTR, several are found in conserved regions and predicted binding sites for RNA binding proteins or miRNAs (Figure 2.1, Appendix B).

To determine which variants would most likely be disease causing mutations, we searched for unique variants found in patients described as having a disease phenotype, but that had no mutations found in the coding region of the *DMD* gene (NODCM patients) (Appendix B). We identified 21 variants in the 3' UTR that were found only in NODCM patients (Appendix B). We filtered this list down to eight variants amongst nine NODCM patients where the variant was located in a conserved region of the 3' UTR (Table 2.1). Four of these patients were from the newborn screening study that collected bloodspots from infants with high creatine kinase (CK) levels (Mendell et al., 2012). All infants with confirmed *DMD* mutations in this study had CK levels over 2000 U/l (Mendell et al., 2012), whereas these four patients in our analysis had CK levels less than 1000 U/l or had no known CK values (Table 2.2), and are likely not dystrophic. The variants in these newborns are likely not disease causing mutations. There is limited clinical information on the remaining five patients who are from diverse populations and

Table 2.1

3’ UTR variants found only in NODCM patients and conserved regions.

Variant	Patient	3' UTR Position	3' UTR Region	# of 3' UTRs
c.*64_*65insC	DOBS060, DC0644	64	Aa	2
c.*571A>G	DOBS0169	571	Ad	1
c.*651G>A	DOBS0154	651	Ma	1
c.*1838_*1841delTGAT	DC0325	1838	Mc	1
c.*2245T>C	DC0399	2245	Da	1
c.*2268C>T	DOBS017	2268	Da	1
c.*2326G>A	DC0585	2326	Da	1
c.*2591T>A	DC0859	2591	Dc	1

Note: The *DMD* 3’ UTR variants identified in our patient population that were found only in patients with no *DMD* coding mutations and in a conserved region are shown. Patients containing these variants are shown. Patients identifiers that start with “DOBS” were patients a newborn screening study (Mendell et al., 2012). Patient identifiers that start with “DC” are patients from clinical referrals. Both the 3’ UTR nucleotide position and region each variant is found in is shown along with the # of 3’ UTRs containing each variant.

Table 2.2

Clinical information for patients with 3’ UTR NODCM variants.

Patient #	Clinical Information
DC0325	Asian Indian; no family history; no DUP (duplication) test
DC0399	African American; Dilated cardiomyopathy; no DUP (duplication) test
DC0585	Hispanic, possible Becker
DC0644	Finnish (Caucasian), cardiomyopathy
DC0859	Hispanic (Puerto Rican), variant near exon 32 (cryptic splice acceptor?)
DOBS0154	Newborn screening bloodspot; CK Value = 782 U/l
DOBS0169	Newborn screening bloodspot; CK Value = 972 U/l
DOBS017	Newborn screening bloodspot; CK Value = unknown
DOBS060	Newborn screening bloodspot; CK Value = 758 U/l

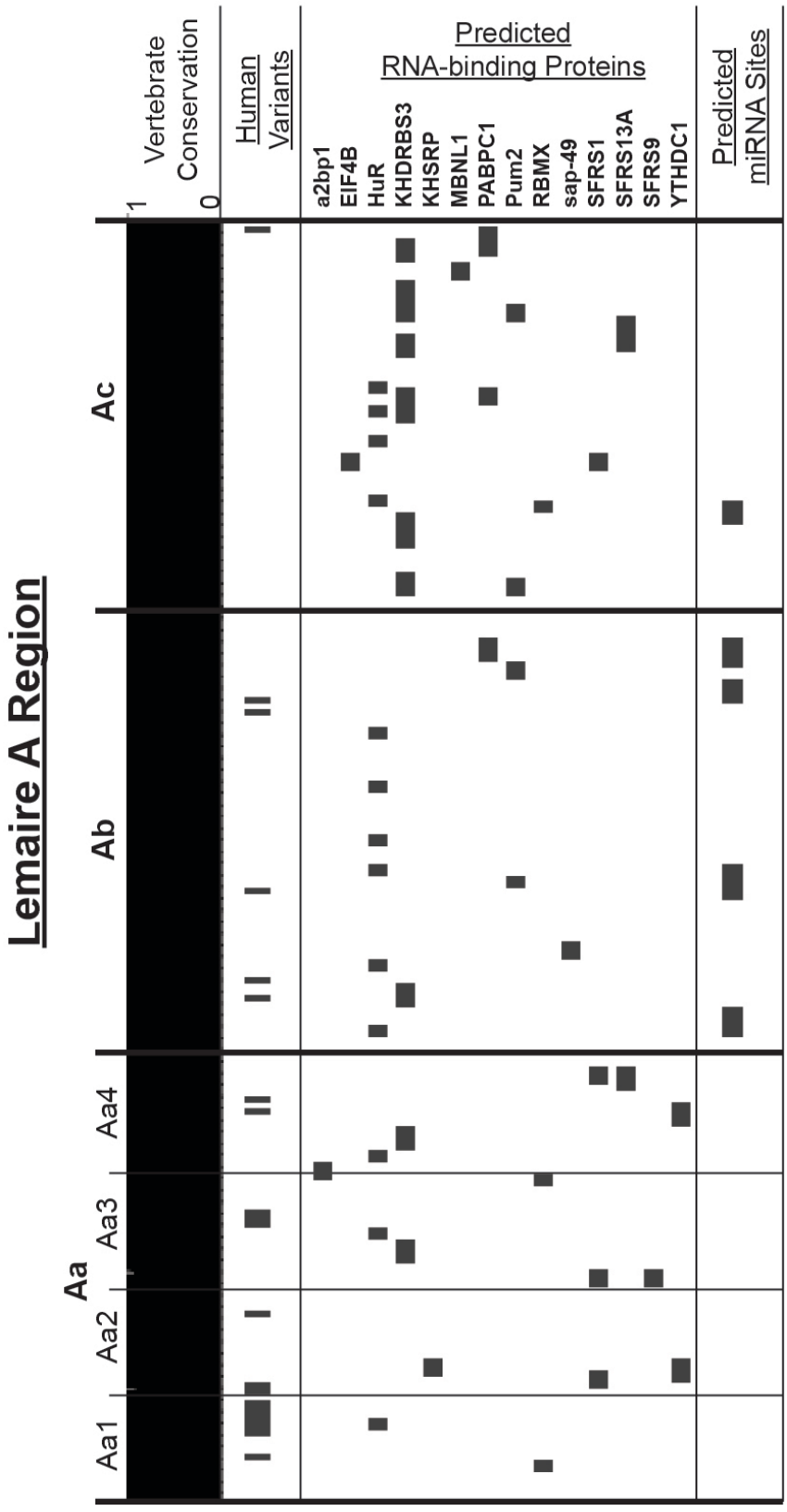
include an Asian Indian, an African American, two Hispanics, and a Caucasian (Table 2.2). Most 3' UTR sequences we obtained were from Caucasians, and these variants could be common in other populations. Testing for duplications using MLPA was not done on two of these patients, and one patient has a possible mutation near exon 32 that might create a cryptic splice acceptor (Table 2.2), evidence that a few of these patients could have disease causing mutations outside of the 3' UTR. We could not identify any clear disease causing mutations from our analysis of human *DMD* 3' UTRs. However, it is possible that variants in the *DMD* 3' UTR could act as modifier mutations that affect disease severity when a second mutation is present. Unfortunately, we could not obtain a big enough cohort of patients with 3' UTR sequences and concise clinical information to investigate whether a human 3' UTR variant is associated with disease severity in DMD and BMD patients. Because of this, we took an experimental approach to better understand the mechanism of how the conserved regions in the 3' UTR function, and to determine which variants or predicted binding sites of miRNAs or RNA binding proteins would most likely effect dystrophin expression and be associated with disease. Due to the large size of conserved regions in the *DMD* 3' UTR, we chose to focus our analysis on the first 150 base pairs of Lemaire A, (called the Aa region in Chapter 1). Lemaire A contains several human variants and predicted binding sites for miRNAs and RNA binding proteins (Figure 2.2, Appendix B) that could potentially impact dystrophin expression when mutated.

The Aa region is sufficient to increase expression in C2C12 cells

Previously, we showed that protein expression decreased by ~70% in C2C12 myotubes when ~400 bp of the Lemaire A region of the *DMD* 3' UTR was deleted. This

Figure 2.2. Variation and predicted binding sites in Lemaire A.

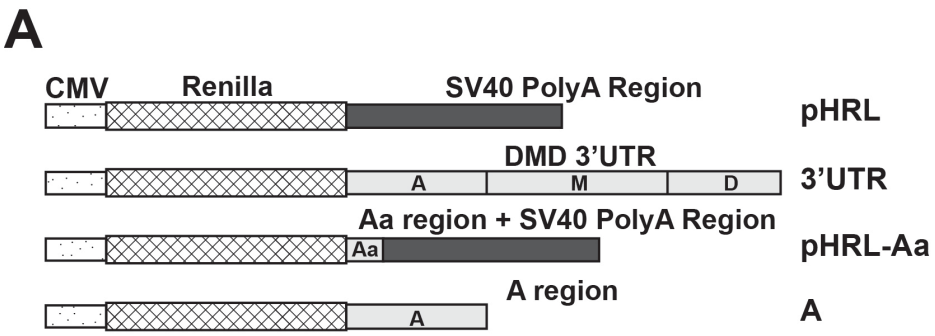
The location of human variants and predicted binding sites for RNA binding proteins and miRNAs in Lemaire A of the *DMD* 3' UTR is shown. Lemaire A was broken down into three regions (Aa, Ab, and Ac), with the Aa region being subdivided into four parts (Aa1, Aa2, Aa3, and Aa4). Human variants were identified in our patient population and in publicly available databases (see Appendix B). Predicted RNA binding protein binding sites were obtained using the RNA Binding Protein DataBase (RBPDB) (Cook et al., 2011). Predicted miRNA binding sites were obtained using TargetScan (Lewis et al., 2005). Vertebrate conservation of Lemaire A was obtained from the Vertebrate PhastCons track from the UCSC genome browser (Kent et al., 2002).

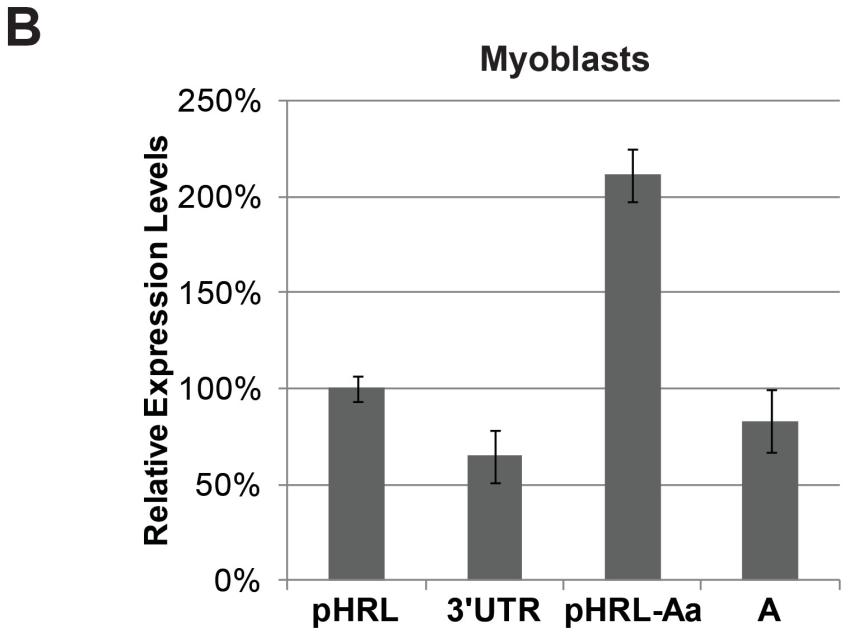

Lemaire A Region
Aa
Aa1
Aa2
Aa3
Aa4
Ab
Ac
1
Vertebrate Conservation
0
Human Variants
Predicted RNA-binding Proteins
a2bp1
EIF4B
HuR
KHDRBS3
KHSRP
MBNL1
PABPC1
Pum2
RBMX
sap-49
SFRS1
SFRS13A
SFRS9
YTHDC1
Predicted miRNA Sites

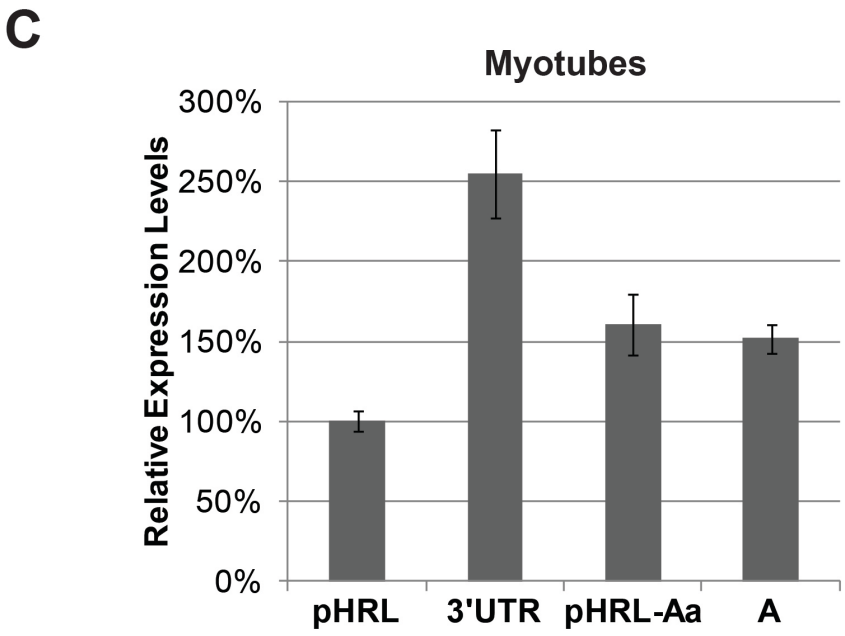
same decrease was observed when only the first 150 nucleotides of Lemaire A was deleted (Aa region) (Chapter 1), showing that this region is necessary for the increase in expression seen during C2C12 differentiation. However, deleting regions downstream had a similar effect on expression and it is unclear whether the Lemaire A region consists of one large regulatory element that must remain intact, or multiple regulatory elements cooperatively regulating expression. To determine whether distinct regulatory elements exist within the sequence of the Aa region of the 3' UTR, and whether it is sufficient to increase expression during differentiation, the Aa region was inserted into the SV40 3' UTR in the pHRL construct (pHRL-Aa, Figure 2.3A), transfected into C2C12 myoblasts and myotubes, and analyzed using a Renilla/Firefly dual luciferase assay (Figure 2.3B and 2.3C, Materials and methods). All transfections were cotransfected with a Firefly construct to normalize for transfection efficiency. Inserting the Aa region into the pHRL construct (pHRL-Aa) increased expression of the construct by ~100% in C2C12 myoblasts, whereas the full length *DMD* 3' UTR (3' UTR) decreased expression by ~45% (Figure 2.3B). In C2C12 myotubes, the Aa region increased expression to 150% compared to the control pHRL construct, whereas the full length 3' UTR construct increased expression to ~250%, showing that the Aa region is sufficient to increase expression in both C2C12 myoblasts and myotubes, but other regions of the 3' UTR suppress or enhance this effect in myoblasts and myotubes, respectively. A pHRL construct containing only the Lemaire A region of the *DMD* 3' UTR (A) only increased expression in C2C12 myotubes (Figure 2.3C). Deleting the Aa region did not affect expression in C2C12 myoblasts (Chapter 1), showing that the increased expression due to the Aa region does not occur in the presence of the full *DMD* 3' UTR and suggests that

Figure 2.3. The Aa region increases expression in C2C12 myoblasts and myotubes.

A. Schematic of the control pHRL-CMV vector (pHRL), the full length *DMD* 3'UTR Renilla construct (3' UTR), the pHRL-CMV vector with the Aa region inserted in the 3' UTR (pHRL-Aa), and the pHRL vector with the Lemaire A region as a 3' UTR (A) is shown. The *DMD* 3' UTR can be divided into three major regions containing the highly conserved Lemaire A and D regions (A and D, respectively) and the less conserved middle region (M). The Aa region consists of the first 150 base pairs of Lemaire A. The pHRL and pHRL-Aa constructs have an intact SV40 polyA region, whereas the 3'UTR and A constructs have an intact *DMD* 3' UTR polyA region.
B. Relative expression of these constructs transfected into C2C12 myoblasts is shown. Expression levels were measured using a dual luciferase assay to measure the relative amounts of Renilla to Firefly protein levels. The results of three biological replicates are shown and normalized to the amount of pHRL expression. Error bars equal +/- 1 standard deviation.
C. Relative expression of the constructs transfected into C2C12 myotubes is shown. The results of three biological replicates are shown and normalized to the amount of pHRL expression. Error bars equal +/- 1 standard deviation.


A
CMV
Renilla
SV40 PolyA Region
pHRL
DMD 3'UTR
A
M
D
3'UTR
Aa region + SV40 PolyA Region
Aa
pHRL-Aa
A region
A
A


B
Myoblasts
Relative Expression Levels
250%
200%
150%
100%
50%
0%
pHRL
3'UTR
pHRL-Aa
A


C
Myotubes
Relative Expression Levels
300%
250%
200%
150%
100%
50%
0%
pHRL
3'UTR
pHRL-Aa
A

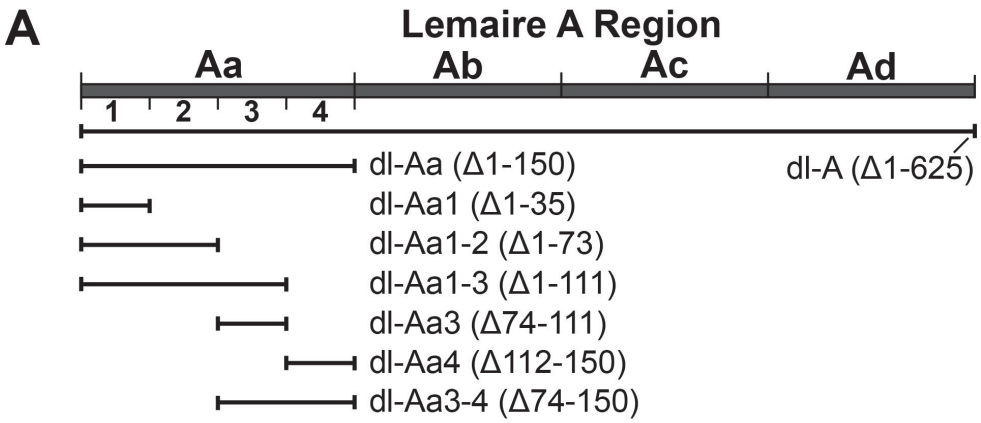
there are additional regions outside the Aa region in the 3' UTR that suppress expression in C2C12 myoblasts but increase expression in myotubes.

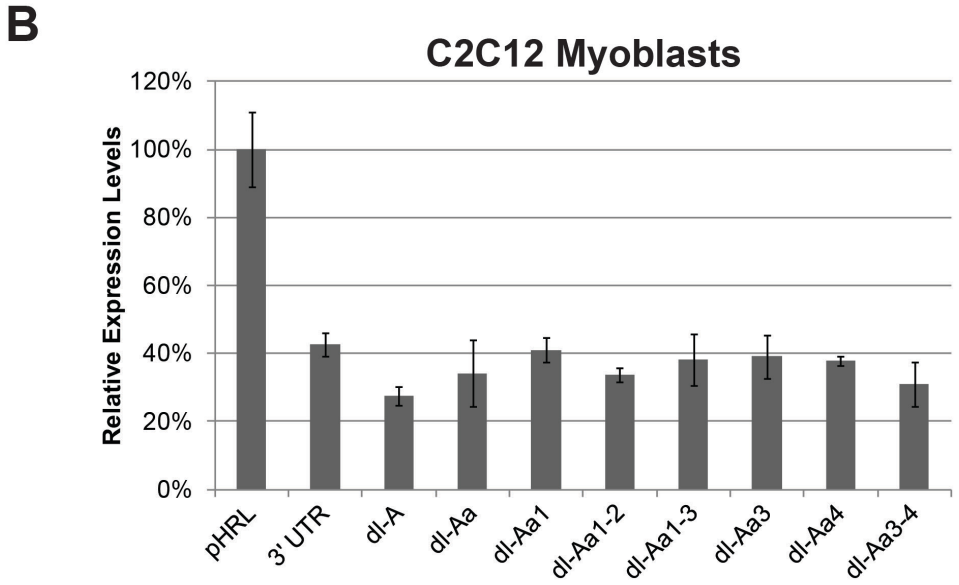
Size and location of deletions in the Aa region affect expression in C2C12 myotubes

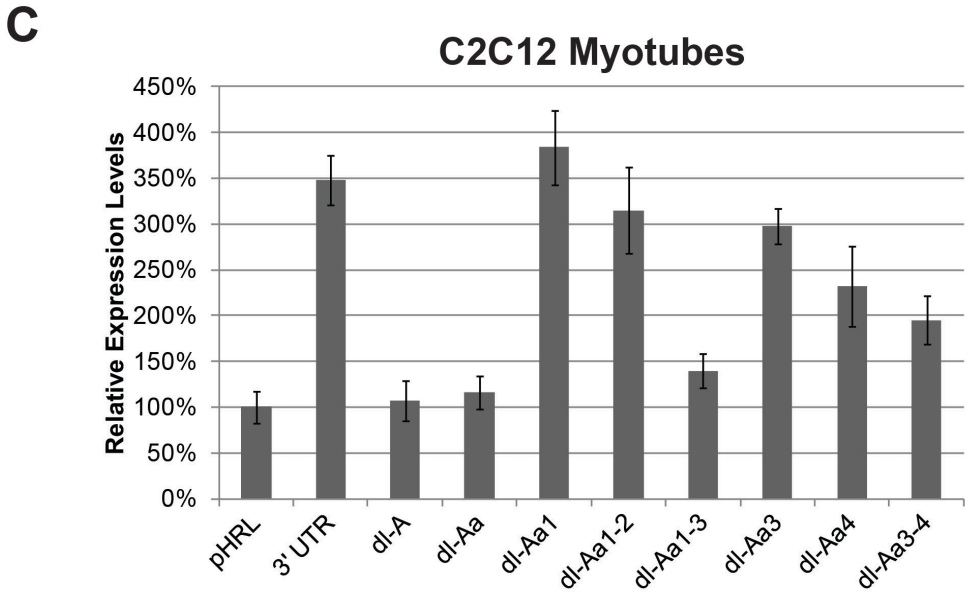
To determine which sequences in the Aa region of the *DMD* 3'UTR were necessary for the increase in expression seen in C2C12 myotubes, the Aa region was subdivided into ~35-40 base pair sections (Aa1, Aa2, Aa3, and Aa4), and 3' UTR deletion constructs were made deleting several variations of these regions (Figure 2.4A). These constructs were transfected into C2C12 myoblasts and myotubes, and expression levels were determined using a Renilla/Firefly dual luciferase assay as previously described. Deleting the entire Aa region or any combination of Aa sections had little impact on expression in C2C12 myoblasts compared to the full length 3' UTR construct (Figure 2.4B), but deleting portions of the Aa region in C2C12 myotubes had variable effects on protein expression (Figure 2.4C). We made deletions of increasing size of the Aa region (Aa1, Aa1-2, and Aa1-3) and found that deleting the Aa1 region (dl-Aa1) had no significant effect on expression, deleting the Aa1 and Aa2 region (dl-Aa1-2) decreased expression by ~15%, and deleting Aa1-3 (dl-Aa1-3) decreased expression by ~60% which was close to the expression levels seen when the entire Lemaire A region was deleted (Figure 2.4C). Interestingly, when the Aa1-2 region or the Aa3 region were deleted separately (dl-Aa1-2 and dl-Aa3), there was very little impact on expression, but when both regions were deleted (dl-Aa1-3), expression decreased by ~60% similar to when the entire Lemaire A region was deleted (dl-A) (Figure 2.4C). Constructs deleting the Aa3 or Aa4 region had the largest impact on expression in C2C12 myotubes,

Figure 2.4. Expression of mini deletion constructs of the Aa region in C2C12 cells.

A. Schematic showing the regions deleted in the Aa region of the *DMD* 3' UTR in several deletion constructs are shown. The *DMD* 3' UTR contains a highly conserved Lemaire A region that was split into ~150 base pair sections (Aa, Ab, Ac, and Ad). The Aa region was further divided into ~35-40 base pair sections (Aa1, Aa2, Aa3, and Aa4). Each bar represents the region deleted in each construct and the numbers correspond to the nucleotide position in the *DMD* 3' UTR.
B. The relative expression of the pHRL construct (pHRL), the full length *DMD* 3' UTR construct (3' UTR), and several deletion constructs transfected into C2C12 myoblasts are shown. Deleting any portion of the Lemaire Aa region had little impact on expression compared to the full length 3' UTR construct (3'UTR). The result of at least three biological replicates is shown with expression levels normalized to pHRL-CMV expression (pHRL) using a Renilla/Firefly dual luciferase assay. Error bars equal +/- 1 standard deviation.
C. The relative expression of the pHRL construct (pHRL), the full length *DMD* 3' UTR construct (3' UTR), and several deletion constructs transfected in C2C12 myotubes are shown. Larger deletions and deletions containing the Aa3 or Aa4 regions had the largest effect on expression. The result of at least three biological replicates is shown with expression levels normalized to pHRL-CMV expression (pHRL) using a Renilla/Firefly dual luciferase assay. Error bars equal +/- 1 standard deviation.


A
Lemaire A Region
Aa
Ab
Ac
Ad
1
2
3
4
dl-Aa (Δ1-150)
dl-A (Δ1-625)
dl-Aa1 (Δ1-35)
dl-Aa1-2 (Δ1-73)
dl-Aa1-3 (Δ1-111)
dl-Aa3 (Δ74-111)
dl-Aa4 (Δ112-150)
dl-Aa3-4 (Δ74-150)


B
C2C12 Myoblasts
Relative Expression Levels
120%
100%
80%
60%
40%
20%
0%
pHRL
3' UTR
dl-A
dl-Aa
dl-Aa1
dl-Aa1-2
dl-Aa1-3
dl-Aa3
dl-Aa4
dl-Aa3-4


C
C2C12 Myotubes
Relative Expression Levels
450%
400%
350%
300%
250%
200%
150%
100%
50%
0%
pHRL
3' UTR
dl-A
dl-Aa
dl-Aa1
dl-Aa1-2
dl-Aa1-3
dl-Aa3
dl-Aa4
dl-Aa3-4

suggesting that these regions affect expression more than the Aa1 and Aa2 regions (Figure 2.4C).

RNA binding proteins bind to the Aa3-4 region of the *DMD* 3' UTR

Because deleting the Aa1-2 region had very little impact on expression, and because the Aa3-4 region had the largest impact on expression in the Aa region, we investigated whether RNA binding proteins were binding to the Aa3-4 region. To do this, [α-^{32}P]UTP labeled RNA of the Aa3-4 region was synthesized and hybridized with whole protein extracts from C2C12 myoblasts or myotubes. Unlabeled tRNA and RNA from the kelch gene in Drosophila were used as competitors to bind to nonspecific RNA binding proteins where indicated. The hybridized RNA/protein complexes were crosslinked using UV light and RNA was digested. The resulting RNase resistant, covalently linked RNA/protein complexes were electrophoresed on an SDS Bis Tris polyacrylamide gel to determine the size of proteins that were binding to the Aa3-4 region (Figure 2.5). Several protein bands were detected with the Aa3-4 RNA when only tRNA was used as a competitor (Figure 2.5, tRNA only). However, when tRNA and the additional competitor RNA were added, we found two strong protein bands that bound to the Aa3-4 region in myotubes (Figure 2.5, tRNA+ Comp. RNA). The unknown binding proteins are ~35-37 kDa and ~80-90 kDa. These bands were not present when a [α-^{32}P]UTP labeled control RNA (kelch RNA) was hybridized with C2C12 cell protein extracts (Figure 2.5, Control RNA).

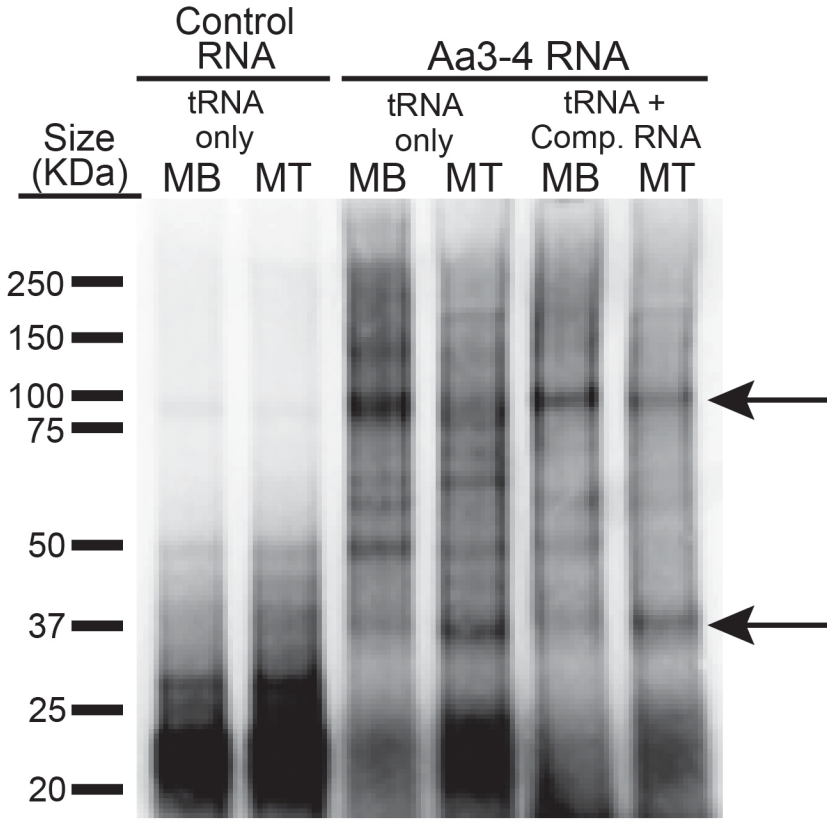


Figure 2.5. Proteins binding to the Aa3-4 region of the *DMD* 3' UTR.

Proteins bound to [α-^{32}P]UTP labeled Aa3-4 RNA (Aa3-4 RNA) or a labeled control RNA (Control RNA) are shown. Labeled RNA was hybridized in either C2C12 myoblast (MB) or myotube (MT) protein extracts. Hybridized RNA protein complexes were crosslinked using UV and RNA was digested. Unlabeled tRNA (tRNA) and RNA from the kelch gene (Comp. RNA) were used as competitor RNA during hybridization. Arrows indicate two strong protein bands at ~35-37 kDa and ~80-90 kDa found when the Aa3-4 RNA was hybridized with C2C12 myotube proteins in the presence of competitor RNA. Proteins were electrophoresed on a precast SDS NuPAGE 4-12% Bis Tris Gel (Life Technologies) and MES running buffer. Protein sizes were determined using the Precision Plus Protein Kaleidoscope™ Standards (Bio Rad).

A human *DMD* 3' UTR variant at a predicted HuR binding site does not alter expression in C2C12 cells

There is only one variant (c.*93_*94InsT) in the Aa region found in our human DMD population that is located at a predicted RNA binding protein binding site. This variant, found in four human DMD patients, has a T inserted at a predicted HuR binding site in the Aa3 region of the *DMD* 3' UTR (Appendix B). To determine whether this variant could alter dystrophin expression in C2C12 cells, we made a full length *DMD* 3' UTR construct containing this variant. We transfected this construct into C2C12 myoblasts and myotubes and measured the effect on expression using a Renilla/Firefly dual luciferase assay as previously described. Inserting a T nucleotide at the predicted HuR binding site (c.*93InsT) had no effect on expression in C2C12 myoblasts or myotubes compared to the wild type 3'UTR construct (3'UTR) (Figure 2.6).

Discussion

Previously, we found that the conserved Lemaire A region of the *DMD* 3' UTR regulates expression by increasing mRNA abundance and translation in differentiated myotubes (Chapter 1). We had previously attempted to narrow down the regulatory sequence using mutagenesis but found that the entire region was necessary for function and deleting any large portion of Lemaire A resulted in loss of this regulation (Chapter 1). This result is consistent with the high level of conservation that spans this region, and is similar to previously described HCEs (Dassi et al., 2013; Siepel et al., 2005). HCEs are enriched in 3' UTRs and may be involved in posttranscriptional gene regulation, alternative splicing, and polyadenylation (Dassi et al., 2013; Ho and Gunderson, 2011; Sathirapongsasuti et al., 2011; Siepel et al., 2005). Although there is evidence that HCEs

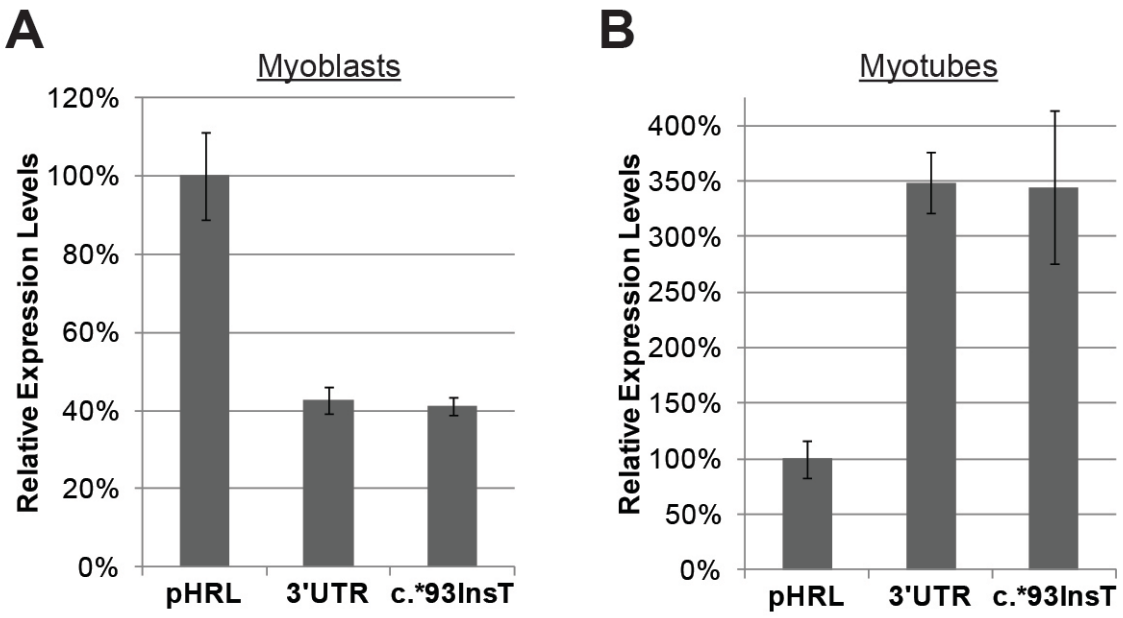


Figure 2.6. The human *DMD* variant c.*93_*94InsT has no effect on expression in C2C12 cells.

A. The relative protein expression of the control pHRL vector (pHRL), a pHRL vector containing the wild type *DMD* 3'UTR (3'UTR), and a mutated construct containing an inserted T nucleotide at a predicted HuR binding site (c.*93InsT) transfected into C2C12 myoblasts are shown. The results of three biological replicates are shown. Error bars equal +/- 1 standard deviation.
B. The relative protein expression in C2C12 myotubes is shown. The results of three biological replicates are shown. Error bars equal +/- 1 standard deviation

interact with RNA binding proteins and miRNAs, the mechanism of how HCEs regulate gene expression is still not understood, and why the extensive length and high conservation of these regions must remain intact is unknown.

Here, we show that the first 150 base pairs of Lemaire A (the Aa region) alone are sufficient to increase expression in both C2C12 myoblasts and myotubes. However, the expression pattern is different than when the entire Lemaire A region is present. The Aa region itself increases expression in both myoblasts and myotubes, whereas the entire *DMD* 3' UTR or a 3' UTR consisting of the Lemaire A region alone decreases expression in myoblasts and only increases expression as the cells differentiate into myotubes. This suggests that the Lemaire A region is not comprised of a single functional regulatory element, but of several functional regulatory elements that may be acting cooperatively to regulate dystrophin expression, and elements contained within the Lemaire A region that lie outside the Aa region suppress increased expression in myoblasts. There is evidence that miRNA-31 binds to the Ac region of Lemaire A and suppresses dystrophin expression (Cacchiarelli et al., 2011). However, our own analysis suggests that there are additional regions that suppress expression in myoblasts (See Chapter 1).

We used deletion mutagenesis to further break down the Aa region into four sections of ~35-40 base pairs (Aa1, Aa2, Aa3, and Aa4) to determine which regions contained regulatory elements. We made deletions of increasing size of the Aa region (Aa1, Aa1-2, and Aa1-3) and found that deleting the Aa1 region (dl-Aa1) had no effect on expression, deleting the Aa1 and Aa2 region (dl-Aa1-2) decreased expression by ~15%, and deleting Aa1-3 (dl-Aa1-3) decreased expression by ~60% and was close to the expression levels seen when the entire Lemaire A region was deleted. We first

hypothesized that the Aa3 region must contain the regulatory element functioning in this region. Unexpectedly, when we deleted only the Aa3 region (dl-Aa3) we did not see this large decrease in expression. This result could be explained by the existence of two redundant regulatory elements contained within the Aa1-2 region and the Aa3 region. Under this scenario, when the Aa1-2 or Aa3 region is deleted separately, the region still retains the ability to increase expression as the second element is present, whereas when both the Aa1-2 and Aa3 regions are deleted, function is lost and expression does not go up. Alternatively, the size of the deletion may be the determining factor on how well this region functions. Perhaps, there is a size threshold for deletions in the Lemaire A region, and once the deleted size hits a certain point, function of the entire region is abolished. This is consistent with our previous attempts to narrow down the functional sequences of the Lemaire A region and found that deleting any portion of Lemaire A (~150 base pairs each) decreased expression to the same degree as deleting the entire region. Although size of deletion does not fully explain our results and some regions appear more important than others. For example, the Aa4 region seems to be more important than the other regions in Aa. When only the Aa4 region is deleted expression decreases by ~45%, and when the Aa3 and Aa4 regions are deleted expression decreases by ~50%. This decrease is larger than the decrease seen when deletions of the same size are made in the Aa1-2 regions.

Using an RNA protein binding assay, we found that two proteins of ~36 kDa and ~80-90 kDa were binding to the Aa3-4 region. There are several predicted RNA protein binding sites in the Aa3-4 region (Figure 2.2, Table 2.3). Of the predicted RNA protein binding sites in this region, the protein YTHDC1 is the only predicted protein of the size

Table 2.3

Predicted RNA protein binding sites in the Aa3-4 region of the *DMD* 3' UTR.

Score	Relative Score	RBP Name	3' UTR Position Start	End	Matching sequence	Protein Size (kDa)
8.647	100%	a2bp1	37	41	GCAUG	~40-43
6.633	93%	SFRS9	1	5	AGGAG	~26
5.629	90%	YTHDC1	55	60	UCAUAC	85, 100
4.774	93%	SFRS13A	64	70	AAAGAGG	31-40
4.620	100%	SFRS1	68	71	AGGA	28
4.620	100%	SFRS1	1	4	AGGA	28
4.604	98%	KHDRBS3	47	52	UAUAAU	38
4.404	100%	ELAVL1	42	45	GUUU	36
4.404	100%	ELAVL1	16	19	GUUU	36
4.403	83%	RBMX	34	37	CCCG	42
4.326	92%	KHDRBS3	9	14	AAUAAA	38
4.142	81%	SFRS13A	66	72	AGAGGAU	31-40

Note: Predicted RNA protein binding sites in the Aa3-4 region of the *DMD* 3' UTR using RBPDB (Cook et al., 2011) are shown. RBPDB predicts binding sites by scoring potential binding sites in the sequence using position weight matrices and assigns a score that is calculated as the sum of the scores of each nucleotide at each position in the position weight matrix. A relative score is also given as a percent of the score to the maximum possible score of the predicted site. Only predicted sites with a relative score above 80% are shown. The 3' UTR start and end positions if each predicted binding site are shown.

range of 80-90 kDa (Table 2.3). Little is known about this RNA binding protein other than that it interacts with the protein emerin (Wilkinson et al., 2003). Mutations in the emerin gene cause Emery Dreifuss muscular dystrophy in humans, and emerin has been shown to function in both muscle and heart tissue (Fairley et al., 1999; Helbling-Leclerc et al., 2002). However, further investigation would be needed to determine if YTHDC1 is interacting with the *DMD* 3' UTR. Of the predicted RNA protein binding sites, there are three proteins (KHDRBS3, SFRS13A, and ELAVL1 (HuR)) that best match the size of the ~36 kDa protein (Table 2.3). KHDRBS3 (a.k.a. T-STAR) is a KH containing RNA binding protein that has been shown to regulate processes in cancer cells (Lei et al., 2011; Lu et al., 2009; Zhang et al., 2006) and may be involved in spermatogenesis

(Venables et al., 1999). The RNA binding protein SFRS13A (a.k.a. SRSF10 and SRp38) has been shown to be involved in sequence specific alternative splicing (Feng et al., 2009; Ling and Estus, 2010; Zhou et al., 2014). However, there is not strong evidence that KHDRBS3 or SFRS13A is functioning in muscle cells. The best candidate protein in the 35-37 kDa range is the protein ELAVL1 (a.k.a. HuR). Two groups using PAR-CLIP, a genome wide method to examine the interaction between RNA binding proteins and mRNA, identified *DMD* as a target for HuR in HeLa and HEK 293 cells (Hafner et al., 2010; Lebedeva et al., 2011; Mukherjee et al., 2011). Although muscle cells were not used in these experiments, HuR has been shown to regulate myogenesis by binding to the 3' UTR of important myogenic factors such as myogenin and MyoD (Deschenes-Furry et al., 2005; Figueroa et al., 2003; van der Giessen et al., 2003; von Roretz et al., 2011). HuR has also been shown to regulate other proteins that increase in expression during differentiating muscle cells (Deschenes-Furry et al., 2005). HuR was shown to directly interact with and regulate acetylcholinesterase, a protein that dramatically increases in expression during differentiation, by binding to an AU rich sequence in the 3' UTR of the acetylcholinesterase transcript (Deschenes-Furry et al., 2005). The regulatory pattern of acetylcholinesterase is very similar to the regulatory pattern of dystrophin. However, further investigation is required to determine whether the unknown 35-35 kDa protein is HuR.

There are several predicted binding sites for RNA binding proteins and miRNAs in the conserved regions of the *DMD* 3' UTR (Figure 2.2, Appendix B), and several of the variants found in human patients lie within or adjacent to these predicted sites. In the Aa region, there is only one variant found in our human DMD population with a variant

at a predicted RNA protein binding site. This variant, c.*93_*94InsT, inserts a T nucleotide at a predicted HuR binding site in the Aa3 region of the *DMD* 3' UTR. This changes the predicted HuR binding site from a GUUU to a GUUUU. When we transfected C2C12 cells with a *DMD* 3' UTR construct containing this mutation, we saw no effect on expression in either myoblasts or myotubes compared to the full length 3' UTR construct. This mutation may not be significant enough to alter HuR binding at this site, or HuR may not be binding to this region of the 3' UTR.

The *DMD* 3' UTR sequences analyzed were from patients enrolled in the United Dystrophinopathy Project (UDP) (Flanigan et al., 2009), a newborn screening study (Mendell et al., 2012), or from clinical referrals. The group with the most clinical certainty of a DMD or BMD diagnosis was the patients included in the UDP study. The patients from clinical referrals were from cases where a doctor diagnosed a dystrophinopathy or suspected dystrophinopathy enough to order *DMD* mutation screenings. Although it is possible that a subset of these patients were misdiagnosed, it is likely that most of these patients were DMD or BMD patients. The majority of the patients from the newborn screening study are not dystrophic patients. We found no candidate disease causing mutations in the *DMD* 3' UTR amongst the patients included in the UDP study where the clinical diagnosis is the most certain. Our top candidate disease causing variants were found only in patients from the newborn screening study and older clinical referrals with limited patient information. Our search for disease causing variants pulled out the patients with incomplete screening with the possibility of mutations in other regions of the gene. We conclude that it is unlikely that a disease causing 3' UTR mutation exists in these patients. However, it is possible that variants

found in the 3' UTR could affect the disease phenotype by acting as modifier mutations when a second mutation is present. Such mutations have been found in other genes and shown to modify disease severity in DMD and BMD patients (Chandrasekharan et al., 2010; Flanigan et al., 2013; Heydemann et al., 2007). It is been shown that stability of the *DMD* transcript has a large impact on dystrophin protein levels in BMD patients (Spitali et al., 2013), and we have shown evidence that the *DMD* 3' UTR can effect mRNA stability and protein expression (Chapter 1, Appendix A). Variants in the 3' UTR could affect dystrophin expression and act as modifier mutations that alter disease severity. Unfortunately, we were unable to obtain a big enough cohort of *DMD* 3' UTR sequences from human patients accompanied with detailed phenotypic data to test this hypothesis.

One aspect that has been underappreciated in clinical DMD studies is the effect of dystrophin expression levels in the brain. The dystrophin glycoprotein complex is involved in several cellular processes in brain development and function (Waite et al., 2012; Waite et al., 2009), and several DMD and BMD patients exhibit neurological disorders. One third of DMD patients have cognitive impairments with the DMD population having a mean full scale intelligence quotient (IQ) score one standard deviation lower than the population average (Cotton et al., 2005). DMD and BMD is also associated with an increased incidence of neurological disorders, such as autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) (Hendriksen and Vles, 2008; Wu et al., 2005). Although mutations in DMD or BMD patients with cognitive defects have been identified in all regions of the *DMD* gene, there is a higher correlation with mutations found in the distal end of the *DMD* gene and cognitive

impairments (Taylor et al., 2010). It is speculated that this is due to mutations that disrupt the dystrophin isoforms Dp140 and Dp71; short dystrophin isoforms transcribed only from the 3' end of the dystrophin gene that are highly expressed in brain tissue (Blake et al., 1992; Lidov et al., 1995; Moizard et al., 1998). It is possible that disease causing mutations exist in the *DMD* 3' UTR but have been overlooked because they cause a nonmuscular phenotype or affect dystrophin isoforms expressed in nonmuscular tissues such as the brain. One such mutation was recently discovered in a patient with a 3 base pair deletion in the coding region of the *DMD* gene that causes intellectual disability, but no muscle phenotype (de Brouwer et al., 2014). Clinical screens for *DMD* mutations have primarily focused on the muscle phenotype of the disease, and, in fact, classification of DMD and BMD patients is determined by the severity of the muscle phenotype (defined by age of ambulatory loss). Although it is very well documented that DMD and BMD patients exhibit cognitive defects, *DMD* mutation screenings have focused primarily on mutations that disrupt the muscle isoform of dystrophin and phenotypic data on cognitive deficits are not as standardized and often not reported in clinical studies, making it difficult to identify these types of mutations in current *DMD* mutation databases.

In summary, we found no mutations in the human *DMD* 3' UTRs that are likely to be disease causing. However, we could not rule out the possibility that variants in the 3' UTR could also act as modifier mutations when a second mutation is present or cause a disease phenotype in nonmuscular tissues such as the brain. We show evidence that the Lemaire A region is composed of multiple regulatory elements that work cooperatively to downregulate dystrophin expression in myoblasts and upregulate expression in myotubes,

and that the Aa region of Lemaire A binds to RNA binding proteins. The entire region is necessary for the increased expression during differentiation, although certain small deletions and minor alterations can be made to the region without affecting expression in C2C12 cells. The high level of conservation that spans this region remains a mystery, and is higher than what would be required to retain individual RNA protein binding sites. I conclude that there must be some unknown component in the function of these regions that lies within the primary sequence that is yet to be discovered.

Materials and methods

Generation of the 3' UTR reporter constructs and mutagenesis

The full length *DMD* 3' UTR was inserted into the pHRL-CMV Renilla vector (Promega) as previously described (Chapter 1) to make the *DMD* 3' UTR reporter construct. The Aa region of the *DMD* 3' UTR (base pair positions 1 through 150), was inserted into the pHRL-CMV Renilla vector (Promega) by first amplifying the Aa region from the *DMD* 3' UTR construct using the primers GCCGCTTTCTAGAAGGAAGTCTTTTCCACATGGC and GCCGCTTTCTAGACTAATCCTCTTTGTTGTATGAATATTATAAAAA that include an overhang sequence containing the XbaI cut site. The resulting PCR product was digested with XbaI (New England Biolabs) and ligated using T4 DNA Ligase (New England Biolabs) as specified by the manufacturer into the XbaI restriction site of the pHRL-CMV vector (Promega) immediately upstream of the SV40 polyadenylation region.

The pHRL construct containing only the Lemaire A region of the *DMD* 3' UTR downstream of the coding sequence was made using the Phusion Site Directed

Mutagenesis Kit (Thermo Scientific) as specified by the manufacturer using the *DMD* 3' UTR construct as a template and the phosphorylated primers GTGTATCTCAATAAAGCACGCAGTTATGTTAC and CCACTCAGCTGACAGTTCTCAAATG to delete the middle region and Lemaire D region of the 3' UTR. The resulting construct contained only the Lemaire A region and an intact *DMD* polyadenylation signal immediately downstream of the Renilla coding sequence.

The construct containing an insertion in a predicted HuR binding site of the *DMD* 3' UTR (c.*93_*94InsT) and the constructs with deletions in the Aa region of the *DMD* 3' UTR were made using the Phusion Site Directed Mutagenesis Kit (Thermo Scientific), as specified by the manufacturer using the *DMD* 3' UTR construct as a template and phosphorylated primers specific for each deletion or the insertion (Table 2.4). All constructs made were sequence verified and QIAGEN kits were used to prepare plasmid DNA.

Cell culture and protein extraction

Mouse C2C12 myoblast cells were purchased from ATCC (CRL-1772). C2C12 myoblasts were cultured in growth medium consisting of Dulbecco's modified Eagle's medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Hyclone) at 37°C and 5% CO_2. Myogenic differentiation was induced by replacing the medium of the C2C12 myoblasts with DMEM/F-12 media supplemented with 2% horse serum (Sigma) for 6 days. Protein extracts for RNA hybridization experiments were obtained by growing C2C12 myoblasts or myotubes in CytoOne T25 Tissue Culture Flasks (USA Scientific), washing three times in a phosphate buffer saline (PBS), and

Table 2.4

Primers used to make deletions in the Lemaire A region.

Construct	Forward Primer	Reverse Primer
dl-A	CATTTGAGAACTGTCAGCTGAGTGG	CCTAGAATTACTGCTCGTTCTTCAGCAC
dl-Aa	AGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAG	CCTAGAATTACTGCTCGTTCTTCAGCAC
dl-Aa1	GCGATGGAGTCCTTAGTATCAGTC	CCTAGAATTACTGCTCGTTCTTCAGCAC
dl-Aa1-2	AGGAGCAGAATAAATGTTTTACAACTC	CCTAGAATTACTGCTCGTTCTTCAGCAC
dl-Aa1-3	ATGGTTTTTATAATATTCATACAACAAAGAG	CCTAGAATTACTGCTCGTTCTTCAGCAC
dl-Aa3	ATGGTTTTTATAATATTCATACAACAAAGAG	TCTTCATCTGTCATGACTGATACTAAG
dl-Aa4	AGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAG	GCGGGAATCAGGAGTTGTAA
dl-Aa3-4	AGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAG	TCTTCATCTGTCATGACTGATACTAAG
c.*93InsT	AGGAGCAGAATAAATGTTTTTACAACTC	TCTTCATCTGTCATGACTGATACTAAG

collecting the cells in twice the volume of the cell pellet in a homogenization buffer (0.3 M Sucrose, 60 mM NaCl, 15 mM Tris, pH 8.0, and 10 mM EDTA). Cells underwent five cycles of freeze thaw to aid in lysis, and centrifuged at 15000 g for 15 minutes at 4°C to remove cell debris. Protein levels were quantified using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific), and protein extracts were stored at -80°C.

Transfections and dual luciferase reporter assay

C2C12 myoblasts ($1x10^4$ cells) were transfected with 25 ng of the pHRL-CMV vector (Promega) or a *DMD* 3' UTR construct using 0.3 uL of Lipofectamine 2000 reagent (Life Technologies) in 96 well half area plates (Corning) with 0.16 cm^2 growth areas. All reactions were cotransfected with 6.25 ng of the pGL4.13 [luc2/SV40] Firefly Luciferase vector (Promega) as a transfection control. To measure protein expression in myoblasts, cells were harvested 48 hours after transfection by adding 14 uL of a passive lysis buffer (Promega). To measure protein levels in myotubes, the growth media for C2C12 myoblasts was replaced 48 hours after transfection with a low serum media to induce differentiation. The differentiated myotubes were harvested after 6 days in differentiation media by adding 14 uL of a passive lysis buffer (Promega). C2C12

myoblasts and myotubes underwent one freeze thaw cycle to aid in lysis. The bioluminescence of Renilla and firefly luciferase was measured using the Dual Luciferase Reporter Assay System (Promega) in a Veritas Microplate Luminometer (Turner Biosystems) to determine luciferase levels, and the Renilla luciferase signal was normalized to the pGL4.13 firefly luciferase signal in each well to account for variation in transfection efficiency. At least three biological replicates were done for each transfection.

Preparation of labeled RNA transcripts

The Aa3-4 region of the *DMD* 3' UTR was amplified using the primers TAATACGACTCACTATAGGGAGGAGCAGAATAAATGTTTTACAACTC and CATGGTTTTTATAATATTCATACAACAAAGAGGATTAG with a T7 promoter contained within the overhang of the forward primer using Taq DNA Polymerase (Sigma Aldrich) and purified using the QIAquick Gel Extraction Kit (Qiagen) using the manufacturer's recommended protocol. Labeled RNA transcripts were synthesized using T7 RNA polymerase and [α-^{32}P]UTP (Perkin Elmer). The *in vitro* transcription reaction had a final concentration of 1 mM of dATP, dCTP, and dGTP, 5 μM dUTP, 40 μCi [α-^{32}P]UTP (Perkin Elmer), 200 ng of DNA, 100 μg/mL Bovine Serum Albumin (BSA) (New England Biolabs), 5 mM DTT, 40 U RNasin® Plus RNase Inhibitor (Promega), and 1 Unit T7 RNA Polymerase in a transcription buffer (60 mM Tris pH 8.0, 50 mM $MgCl_2$, 4 mM spermadine, 0.02% Triton X-100 (Sigma Aldrich)). The reaction was incubated for 45 minutes at 37°C before 2 U RNase free DNAse (Thermo Scientific) was added and incubated for an additional 10 minutes. The labeled RNA was purified using NucAway Spin Columns (Ambion) as specified by the manufacturer. Labeled and

unlabeled RNA from the kelch gene in Drosophila was obtained from Norma Wells (University of Utah, Human Genetics Department), and used as control RNA and competitor RNA in the RNA protein hybridizations.

RNA protein binding

Labeled RNA from the Aa3-4 region of the *DMD* 3' UTR or a control RNA from the kelch gene in drosophila was hybridized with C2C12 protein extracts by adding 1.5 x 10^5 CPM of labeled RNA with 30 µg of protein from either C2C12 myoblasts or myotubes. Four µg of tRNA was added to each reaction to compete with nonspecific RNA binding proteins, and 600 ng of unlabeled kelch RNA was added as competitor RNA when specified. Each reaction contained 40 units RNasin® Plus RNase Inhibitor (Promega) and a homogenization buffer (0.3 M Sucrose, 60 mM NaCl, 15 mM Tris, pH 8.0, and 10 mM EDTA). Reactions were incubated for 20 minutes at room temperature. The reactions underwent UV crosslinking for 15 minutes at 254 nm UV light using the UV Stratalinker™ 2400 (Stratagene) as specified by the manufacturer. The reaction was digested using 20 ug of RNase A (Life Technologies) and incubating at 37°C for 10 minutes. Electrophoresis was done on an SDS gel using the precast NuPAGE 4-12% Bis Tris Gel (Life Technologies) and MES running buffer as specified by the manufacturer. Protein sizes were determined using the Precision Plus Protein Kaleidoscope™ Standards (Bio Rad). The gel was dried on Whatman paper for 1.5 hours at 80°C using the Model 583 Gel Dryer (Bio Rad), and analyzed using the Typhoon Trio+ Imager (GE Healthcare Life Sciences) and ImageQuant TL Software (GE Healthcare Life Sciences).

Analysis of human *DMD* variants

DMD 3' UTR sequences and clinical information for human DMD patients were obtained from Dr. Robert Weiss (University of Utah, Department of human genetics). Any sequence with a discrepancy that varied from the reference *DMD* sequence was visually inspected using Consed (Gordon and Green, 2013) to determine the validity of each discrepancy. Identified variants were named using the HGVS guidelines and recommendations for variant nomenclature (den Dunnen and Antonarakis, 2000).

Acknowledgements

I thank Dr. Michael T. Howard (University of Utah, Human Genetics Department) for his guidance and suggestions on this work. I thank Dr. Robert Weiss (University of Utah, Human Genetics Department) for contributing human 3' UTR sequencing data, and for his guidance and suggestions on this work. I thank Chris Anderson and Norma Wills for technical assistance.

References

Aartsma-Rus, A., Van Deutekom, J.C., Fokkema, I.F., Van Ommen, G.J., and Den Dunnen, J.T. (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve *34*, 135-144.

Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.R., Gunel, M.*, et al.* (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science *310*, 317-320.

Allamand, V., Richard, P., Lescure, A., Ledeuil, C., Desjardin, D., Petit, N., Gartioux, C., Ferreiro, A., Krol, A., Pellegrini, N.*, et al.* (2006). A single homozygous point mutation in a 3'untranslated region motif of selenoprotein N mRNA causes SEPN1-related

myopathy. EMBO Rep *7*, 450-454.

Andreassi, C., and Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. Trends Cell Biol *19*, 465-474.

Beffagna, G., Occhi, G., Nava, A., Vitiello, L., Ditadi, A., Basso, C., Bauce, B., Carraro, G., Thiene, G., Towbin, J.A.*, et al.* (2005). Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. Cardiovasc Res *65*, 366-373.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science *304*, 1321-1325.

Blake, D.J., Love, D.R., Tinsley, J., Morris, G.E., Turley, H., Gatter, K., Dickson, G., Edwards, Y.H., and Davies, K.E. (1992). Characterization of a 4.8kb transcript from the Duchenne muscular dystrophy locus expressed in Schwannoma cells. Hum Mol Genet *1*, 103-109.

Cacchiarelli, D., Incitti, T., Martone, J., Cesana, M., Cazzella, V., Santini, T., Sthandier, O., and Bozzoni, I. (2011). miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy. EMBO Rep *12*, 136-141.

Calin, G.A., Liu, C.G., Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., Fabbri, M., Cimmino, A., Lee, E.J., Wojcik, S.E.*, et al.* (2007). Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell *12*, 215-229.

Chandrasekharan, K., Yoon, J.H., Xu, Y., deVries, S., Camboni, M., Janssen, P.M., Varki, A., and Martin, P.T. (2010). A human-specific deletion in mouse Cmah increases disease severity in the mdx model of Duchenne muscular dystrophy. Sci Transl Med *2*, 42ra54.

Chen, J.M., Ferec, C., and Cooper, D.N. (2006a). A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. Hum Genet *120*, 1-21.

Chen, J.M., Ferec, C., and Cooper, D.N. (2006b). A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. Hum Genet *120*, 301-333.

Cirak, S., Arechavala-Gomeza, V., Guglieri, M., Feng, L., Torelli, S., Anthony, K., Abbs, S., Garralda, M.E., Bourke, J., Wells, D.J.*, et al.* (2011). Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. Lancet *378*, 595-605.

Cook, K.B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T.R. (2011). RBPDB: a database of RNA-binding specificities. Nucleic Acids Res *39*, D301-308.

Cotton, S.M., Voudouris, N.J., and Greenwood, K.M. (2005). Association between intellectual functioning and age in children and young adults with Duchenne muscular dystrophy: further results from a meta-analysis. Dev Med Child Neurol *47*, 257-265.

Dassi, E., Zuccotti, P., Leo, S., Provenzani, A., Assfalg, M., D'Onofrio, M., Riva, P., and Quattrone, A. (2013). Hyper conserved elements in vertebrate mRNA 3'-UTRs reveal a translational network of RNA-binding proteins controlled by HuR. Nucleic Acids Res *41*, 3201-3216.

de Brouwer, A.P., Nabuurs, S.B., Verhaart, I.E., Oudakker, A.R., Hordijk, R., Yntema, H.G., Hordijk-Hos, J.M., Voesenek, K., de Vries, B.B., van Essen, T.*, et al.* (2014). A 3-base pair deletion, c.9711_9713del, in DMD results in intellectual disability without muscular dystrophy. Eur J Hum Genet *22*, 480-485.

den Dunnen, J.T., and Antonarakis, S.E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat *15*, 7-12.

Deschenes-Furry, J., Belanger, G., Mwanjewe, J., Lunde, J.A., Parks, R.J., Perrone-Bizzozero, N., and Jasmin, B.J. (2005). The RNA-binding protein HuR binds to acetylcholinesterase transcripts and regulates their expression in differentiating skeletal muscle cells. J Biol Chem *280*, 25361-25368.

Fairclough, R.J., Wood, M.J., and Davies, K.E. (2013). Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches. Nat Rev Genet *14*, 373-378.

Fairley, E.A., Kendrick-Jones, J., and Ellis, J.A. (1999). The Emery-Dreifuss muscular dystrophy phenotype arises from aberrant targeting and binding of emerin at the inner nuclear membrane. J Cell Sci *112 (Pt 15)*, 2571-2582.

Feng, Y., Valley, M.T., Lazar, J., Yang, A.L., Bronson, R.T., Firestein, S., Coetzee, W.A., and Manley, J.L. (2009). SRp38 regulates alternative splicing and is required for Ca(2+) handling in the embryonic heart. Dev Cell *16*, 528-538.

Figueroa, A., Cuadrado, A., Fan, J., Atasoy, U., Muscat, G.E., Munoz-Canoves, P., Gorospe, M., and Munoz, A. (2003). Role of HuR in skeletal myogenesis through coordinate regulation of muscle differentiation genes. Mol Cell Biol *23*, 4991-5004.

Flanigan, K.M., Dunn, D.M., von Niederhausern, A., Soltanzadeh, P., Gappmaier, E., Howard, M.T., Sampson, J.B., Mendell, J.R., Wall, C., King, W.M.*, et al.* (2009). Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. Hum Mutat *30*, 1657-1666.

Flanigan, K.M., Ceco, E., Lamar, K.M., Kaminoh, Y., Dunn, D.M., Mendell, J.R., King,

W.M., Pestronk, A., Florence, J.M., Mathews, K.D., *et al.* (2013). LTBP4 genotype predicts age of ambulatory loss in duchenne muscular dystrophy. Ann Neurol *73*, 481-488.

Goemans, N.M., Tulinius, M., van den Akker, J.T., Burm, B.E., Ekhart, P.F., Heuvelmans, N., Holling, T., Janson, A.A., Platenburg, G.J., Sipkens, J.A., *et al.* (2011). Systemic administration of PRO051 in Duchenne's muscular dystrophy. N Engl J Med *364*, 1513-1522.

Gordon, D., and Green, P. (2013). Consed: a graphical editor for next-generation sequencing. Bioinformatics *29*, 2936-2937.

Gramolini, A.O., Belanger, G., and Jasmin, B.J. (2001). Distinct regions in the 3' untranslated region are responsible for targeting and stabilizing utrophin transcripts in skeletal muscle cells. J Cell Biol *154*, 1173-1183.

Greener, M.J., Sewry, C.A., Muntoni, F., and Roberts, R.G. (2002). The 3'-untranslated region of the dystrophin gene - conservation and consequences of loss. Eur J Hum Genet *10*, 413-420.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.C., Munschauer, M., *et al.* (2010). PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. J Vis Exp.

Helbling-Leclerc, A., Bonne, G., and Schwartz, K. (2002). Emery-Dreifuss muscular dystrophy. Eur J Hum Genet *10*, 157-161.

Hendriksen, J.G., and Vles, J.S. (2008). Neuropsychiatric disorders in males with duchenne muscular dystrophy: frequency rate of attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder, and obsessive--compulsive disorder. J Child Neurol *23*, 477-481.

Heydemann, A., Doherty, K.R., and McNally, E.M. (2007). Genetic modifiers of muscular dystrophy: implications for therapy. Biochim Biophys Acta *1772*, 216-228.

Ho, E.S., and Gunderson, S.I. (2011). Long conserved fragments upstream of Mammalian polyadenylation sites. Genome Biol Evol *3*, 654-666.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Kinali, M., Arechavala-Gomeza, V., Feng, L., Cirak, S., Hunt, D., Adkin, C., Guglieri, M., Ashton, E., Abbs, S., Nihoyannopoulos, P., *et al.* (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study.

Lancet Neurol *8*, 918-928.

Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C., and Kunkel, L.M. (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell *50*, 509-517.

Lebedeva, S., Jens, M., Theil, K., Schwanhausser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Mol Cell *43*, 340-352.

Lei, K.F., Liu, B.Y., Wang, Y.F., Chen, X.H., Yu, B.Q., Guo, Y., and Zhu, Z.G. (2011). SerpinB5 interacts with KHDRBS3 and FBXO32 in gastric cancer cells. Oncol Rep *26*, 1115-1120.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Lidov, H.G., Selig, S., and Kunkel, L.M. (1995). Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. Hum Mol Genet *4*, 329-335.

Ling, I.F., and Estus, S. (2010). Role of SFRS13A in low-density lipoprotein receptor splicing. Hum Mutat *31*, 702-709.

Lu, Y., Ryan, S.L., Elliott, D.J., Bignell, G.R., Futreal, P.A., Ellison, D.W., Bailey, S., and Clifford, S.C. (2009). Amplification and overexpression of Hsa-miR-30b, Hsa-miR-30d and KHDRBS3 at 8q24.22-q24.23 in medulloblastoma. PLoS One *4*, e6159.

Matoulkova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol *9*, 563-576.

Mazumder, B., Seshadri, V., and Fox, P.L. (2003). Translational control by the 3'-UTR: the ends specify the means. Trends Biochem Sci *28*, 91-98.

McCabe, E.R., Towbin, J., Chamberlain, J., Baumbach, L., Witkowski, J., van Ommen, G.J., Koenig, M., Kunkel, L.M., and Seltzer, W.K. (1989). Complementary DNA probes for the Duchenne muscular dystrophy locus demonstrate a previously undetectable deletion in a patient with dystrophic myopathy, glycerol kinase deficiency, and congenital adrenal hypoplasia. J Clin Invest *83*, 95-99.

Mendell, J.R., Shilling, C., Leslie, N.D., Flanigan, K.M., al-Dahhak, R., Gastier-Foster, J., Kneile, K., Dunn, D.M., Duval, B., Aoyagi, A.*, et al.* (2012). Evidence-based path to newborn screening for Duchenne muscular dystrophy. Ann Neurol *71*, 304-313.

Moizard, M.P., Billard, C., Toutain, A., Berret, F., Marmin, N., and Moraine, C. (1998). Are Dp71 and Dp140 brain dystrophin isoforms related to cognitive impairment in Duchenne muscular dystrophy? Am J Med Genet *80*, 32-41.

Muir, L.A., and Chamberlain, J.S. (2009). Emerging strategies for cell and gene therapy of the muscular dystrophies. Expert Rev Mol Med *11*, e18.

Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Jr., Tuschl, T., Ohler, U., and Keene, J.D. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. Mol Cell *43*, 327-339.

Muntoni, F., Torelli, S., and Ferlini, A. (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol *2*, 731-740.

Pichavant, C., Aartsma-Rus, A., Clemens, P.R., Davies, K.E., Dickson, G., Takeda, S., Wilton, S.D., Wolff, J.A., Wooddell, C.I., Xiao, X.*, et al.* (2011). Current status of pharmaceutical and genetic therapeutic approaches to treat DMD. Mol Ther *19*, 830-840.

Pillers, D.A., Towbin, J.A., Chamberlain, J.S., Wu, D., Ranier, J., Powell, B.R., and McCabe, E.R. (1990). Deletion mapping of Aland Island eye disease to Xp21 between DXS67 (B24) and Duchenne muscular dystrophy. Am J Hum Genet *47*, 795-801.

Sathirapongsasuti, J.F., Sathira, N., Suzuki, Y., Huttenhower, C., and Sugano, S. (2011). Ultraconserved cDNA segments in the human transcriptome exhibit resistance to folding and implicate function in translation and alternative splicing. Nucleic Acids Res *39*, 1967-1979.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res *29*, 308-311.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S.*, et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.

Spitali, P., van den Bergen, J.C., Verhaart, I.E., Wokke, B., Janson, A.A., van den Eijnde, R., den Dunnen, J.T., Laros, J.F., Verschuuren, J.J., t Hoen, P.A.*, et al.* (2013). DMD transcript imbalance determines dystrophin levels. FASEB J *27*, 4909-4916.

Taylor, P.J., Betts, G.A., Maroulis, S., Gilissen, C., Pedersen, R.L., Mowat, D.R., Johnston, H.M., and Buckley, M.F. (2010). Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy. PLoS One *5*, e8803.

Todorova, A., Todorov, T., Georgieva, B., Lukova, M., Guergueltcheva, V., Kremensky,

I., and Mitev, V. (2008). MLPA analysis/complete sequencing of the DMD gene in a group of Bulgarian Duchenne/Becker muscular dystrophy patients. Neuromuscul Disord *18*, 667-670.

van der Giessen, K., Di-Marco, S., Clair, E., and Gallouzi, I.E. (2003). RNAi-mediated HuR depletion leads to the inhibition of muscle cell differentiation. J Biol Chem *278*, 47119-47128.

van Deutekom, J.C., Janson, A.A., Ginjaar, I.B., Frankhuizen, W.S., Aartsma-Rus, A., Bremmer-Bout, M., den Dunnen, J.T., Koop, K., van der Kooi, A.J., Goemans, N.M.*, et al.* (2007). Local dystrophin restoration with antisense oligonucleotide PRO051. N Engl J Med *357*, 2677-2686.

van Putten, M., Hulsker, M., Nadarajah, V.D., van Heiningen, S.H., van Huizen, E., van Iterson, M., Admiraal, P., Messemaker, T., den Dunnen, J.T., t Hoen, P.A.*, et al.* (2012). The effects of low levels of dystrophin on mouse muscle function and pathology. PLoS One *7*, e31937.

van Putten, M., Hulsker, M., Young, C., Nadarajah, V.D., Heemskerk, H., van der Weerd, L., t Hoen, P.A., van Ommen, G.J., and Aartsma-Rus, A.M. (2013). Low dystrophin levels increase survival and improve muscle pathology and function in dystrophin/utrophin double-knockout mice. FASEB J *27*, 2484-2495.

Venables, J.P., Vernet, C., Chew, S.L., Elliott, D.J., Cowmeadow, R.B., Wu, J., Cooke, H.J., Artzt, K., and Eperon, I.C. (1999). T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis. Hum Mol Genet *8*, 959-969.

von Roretz, C., Beauchamp, P., Di Marco, S., and Gallouzi, I.E. (2011). HuR and myogenesis: being in the right place at the right time. Biochim Biophys Acta *1813*, 1663-1667.

Waite, A., Tinsley, C.L., Locke, M., and Blake, D.J. (2009). The neurobiology of the dystrophin-associated glycoprotein complex. Ann Med *41*, 344-359.

Waite, A., Brown, S.C., and Blake, D.J. (2012). The dystrophin-glycoprotein complex in brain development and disease. Trends Neurosci *35*, 487-496.

Wilkinson, F.L., Holaska, J.M., Zhang, Z., Sharma, A., Manilal, S., Holt, I., Stamm, S., Wilson, K.L., and Morris, G.E. (2003). Emerin interacts in vitro with the splicing-associated factor, YT521-B. Eur J Biochem *270*, 2459-2466.

Wu, J.Y., Kuban, K.C., Allred, E., Shapiro, F., and Darras, B.T. (2005). Association of Duchenne muscular dystrophy with autism spectrum disorder. J Child Neurol *20*, 790-795.

Zhang, L., Guo, L., Peng, Y., and Chen, B. (2006). Expression of T-STAR gene is associated with regulation of telomerase activity in human colon cancer cell line HCT-116. World J Gastroenterol *12*, 4056-4060.

Zhou, X., Wu, W., Li, H., Cheng, Y., Wei, N., Zong, J., Feng, X., Xie, Z., Chen, D., Manley, J.L.*, et al.* (2014). Transcriptome analysis of alternative splicing events regulated by SRSF10 reveals position-dependent splicing modulation. Nucleic Acids Res *42*, 4019-4030.

CHAPTER 3

CANINE MODELS OF DUCHENNE MUSCULAR DYSTROPHY

Abstract

Dogs affected by canine X linked muscular dystrophy (*cxmd*) are important animal models of Duchenne muscular dystrophy (DMD), in large part because they recapitulate skeletal muscle pathology to a greater degree than the standard *mdx* mouse model. We studied four dystrophic dog lines that were identified in clinical veterinary practice due to variable weakness and either altered or absent dystrophin expression and we identified novel *DMD* mutations in three of them. These include a Labrador retriever, in which we identified a disease associated LINE-1 insertion resulting in the inclusion of a pseudoexon within the mRNA; a Cocker spaniel, in which we identified a 4 nucleotide frameshifting deletion in exon 65; and a Tibetan terrier, in which we identified a large deletion of exons 8-29. To aid in establishing and maintaining research colonies, genotyping assays were developed for each dog model along with a genotyping assay for a Yorkshire terrier with a known exon 3-7 deletion of the *DMD* gene. Developing therapeutic approaches to treat human DMD patients target specific types of mutations, and the identification of these novel mutations adds to the catalog of available dog lines that may facilitate preclinical studies directed toward these specific mutation classes.

Introduction

The dystrophinopathies Duchenne muscular dystrophy (DMD) and the milder disorder Becker muscular dystrophy (BMD) are both caused by mutations in the *DMD* gene. The *DMD* gene is the largest known human gene, spanning 2.2 megabases on the X chromosome, and consisting of 79 exons. *DMD* codes for several isoforms of the protein dystrophin derived from tissue specific promoters found within the gene (Hoffman et al., 1987; Muntoni et al., 2003). The predominant muscle isoform, Dp427m, codes for a 427 kilodalton protein that connects the cytoskeleton of muscle to the extracellular matrix via the dystrophin associated glycoprotein complex (DAG), and mediates signal transduction cascades through the C terminal domain of dystrophin (Blake et al., 2002; Cohn and Campbell, 2000). Patients with DMD exhibit severe muscle degeneration, development of cardiac hypertrophy, an association with neurological disorders, and a life expectancy into the second decade of life. BMD patients, on the other hand, exhibit muscle degeneration that occurs at a much slower rate and typically survive into late adulthood. The molecular pathogenesis of DMD is not completely understood, but, in the absence of dystrophin, the integrity of the muscle membrane is compromised during muscle contraction, and degeneration of myofibers follows (Deconinck and Dan, 2007).

DMD homologues have been identified in several mammalian and nonmammalian species such as the dog (Cooper et al., 1988; Valentine et al., 1988), cat (Carpenter et al., 1989), mouse (Bulfield et al., 1984), zebrafish (Bassett and Currie, 2004), and *C. elegans* (Bessou et al., 1998), and animal models with mutations in the *DMD* homologue have been used to study Duchenne muscular dystrophy. The most

commonly used animal model is the *mdx* mouse (C57BL/10ScSn-Dmd^{mdx}/J), a strain that contains a nonsense mutation in exon 23 of the *DMD* gene and has no dystrophin expression (Bulfield et al., 1984; Sicinski et al., 1989). It is not a perfect model of human DMD. This is in part because the severe muscle degeneration and early death observed in human DMD patients is not seen in the *mdx* mouse despite the absence of dystrophin expression, and, instead, a relatively mild skeletal muscle phenotype is seen (Collins and Morgan, 2003; Dangain and Vrbova, 1984). Likewise, nonmammalian models of DMD, such as the zebrafish and C. elegans, have the advantage of being easier to genetically manipulate and maintain in large numbers, but have different musculature and pathology compared to human patients (Baumeister and Ge, 2002; Bessou et al., 1998; Chambers et al., 2001; Guyon et al., 2003).

Canine X linked muscular dystrophy (*cxmd*) dogs with mutations in the *DMD* gene have been identified that have an absence of dystrophin expression, and similar clinical features and disease progression as human DMD patients (Valentine et al., 1988). The best characterized dog model to date is the Golden Retriever Muscular Dystrophy (GRMD) dog (Cooper et al., 1988; Kornegay et al., 1988; Valentine et al., 1988). The GRMD dog has a point mutation in the splice acceptor site of intron 6, and, as a result, exon 7 is skipped during splicing of the premRNA, resulting in an out-of-frame transcript and a premature stop codon (Sharp et al., 1992). Other known sporadic canine models of DMD have been found in other breeds. For example, a German short hair pointer with a complete deletion of the *DMD* gene (Schatzberg et al., 1999), a Cavalier King Charles Spaniel with a missense mutation in the 5' donor splice site of exon 50 that results in the deletion of exon 50 in the *DMD* mRNA (Walmsley et al., 2010), and a Pembroke Welsh

corgi with a long interspersed repetitive element-1 (LINE-1) insertion in intron 13 that introduces a pseudoexon in the *DMD* mRNA (Smith et al., 2011) have been found. Several other dystrophin deficient dog lines have been identified, but specific mutations have not been reported (Baltzer et al., 2007; Baroncelli et al., 2013; Barthelemy et al., 2007; Bergman et al., 2002; Blot et al., 2002; Jones et al., 2004; Klarenbeek et al., 2007; Wetterman et al., 2000; Wieczorek et al., 2006).

Several therapeutic approaches are being developed to treat DMD patients (Fairclough et al., 2013; Muir and Chamberlain, 2009; Pichavant et al., 2011), including exon skipping using antisense molecules (Cirak et al., 2011; Goemans et al., 2011; Gurvich et al., 2008; Kinali et al., 2009; Lu et al., 2005; Mendell et al., 2013; van Deutekom et al., 2007), premature stop codon suppression therapies (Barton-Davis et al., 1999; Finkel, 2010; Howard et al., 2004; Malik et al., 2010) and the delivery and expression of mini dystrophin gene constructs lacking nonessential *DMD* coding exons (Fabb et al., 2002; Foster et al., 2012; Foster et al., 2008; Koo et al., 2011a; Koo et al., 2011b; Rodino-Klapac et al., 2010; Wang et al., 2000; Wang et al., 2007; Wang et al., 2012). Animal models have proven to be important tools in testing these developing treatments for human use. Although animal models, such as the *mdx* mouse and nonmammalian models, have been useful in understanding DMD pathogenesis, differences in the musculature and pathological expression of the disease in these animals prevent them from being perfect animal models of DMD when testing these treatments in human DMD patients (Collins and Morgan, 2003; Willmann et al., 2009). Large animal models that more faithfully recapitulate human DMD are particularly useful for preclinical development of potential therapies to better study the effectiveness and side

effects in human patients.

We set out to identify *DMD* mutations in dogs who presented in clinical veterinary practice with symptoms of weakness along with altered or absent dystrophin expression on muscle biopsy. We identified the mutations responsible in three such dogs. These include a Labrador retriever with a LINE-1 insertion in intron 19 of the *DMD* gene that results in the inclusion of a pseudoexon in the mRNA; a Tibetan terrier with a deletion of exons 8-29 of the *DMD* gene; and a Cocker Spaniel with a small frameshifting deletion of 4 nucleotides in exon 65 of the *DMD* gene. We developed genotyping assays for each mutation found to aid in establishing research colonies of these dogs, including a genotyping assay for a Yorkshire terrier with a known exon 3 through 7 deletion of the *DMD* gene where no rapid assay previously existed. Identification of these novel mutations adds to the catalog of available dog models, and may facilitate preclinical studies directed toward these specific mutation classes.

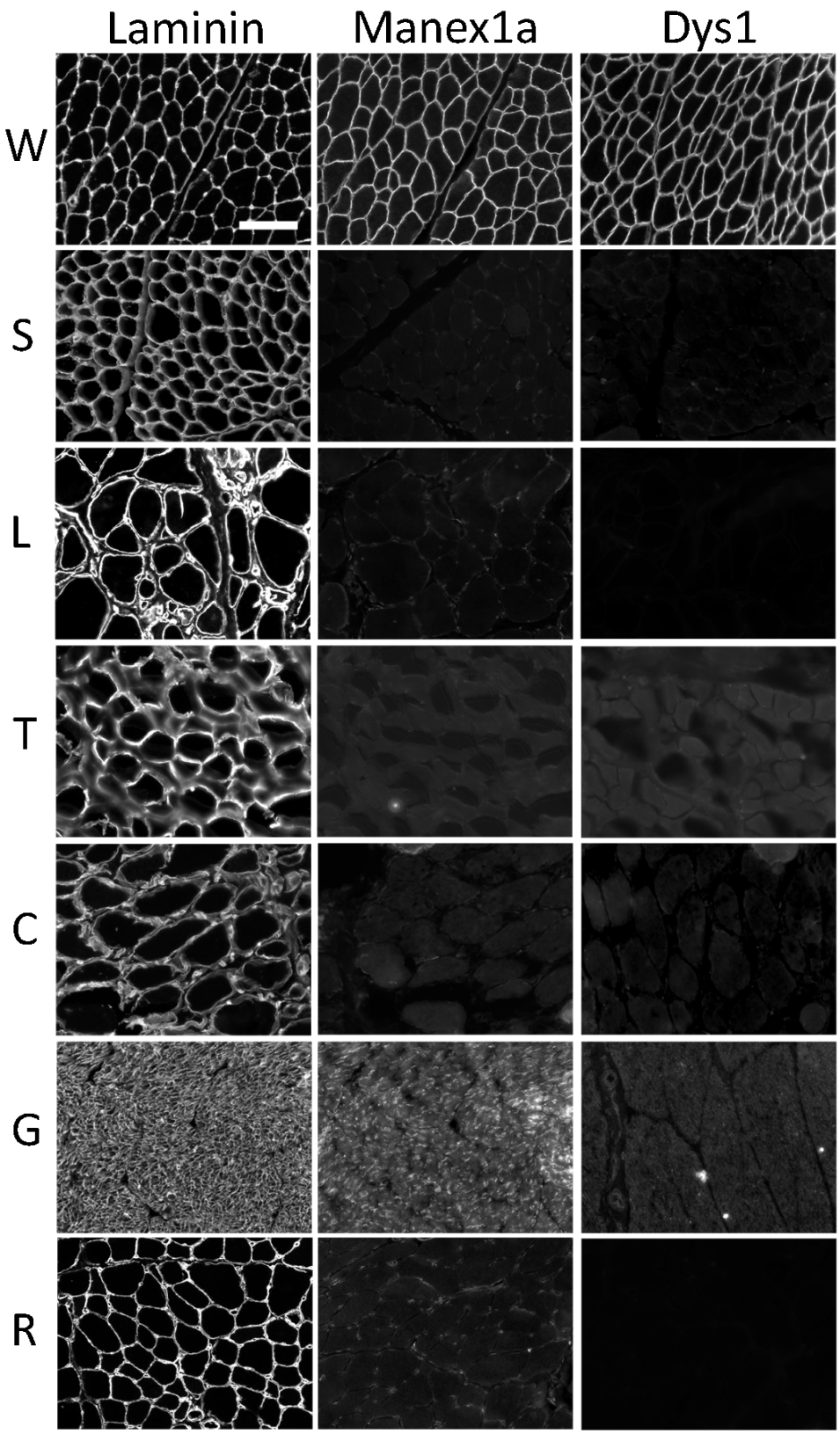
Results

Immunofluorescence/immunoblot analysis of dystrophic dogs

To determine dystrophin expression levels in each dog, muscle sections were obtained from a biopsy of a dystrophic Springer spaniel, Labrador retriever, Tibetan terrier, and Cocker spaniel dogs that had not been screened for *DMD* mutations. Sections of each muscle sample were stained for dystrophin using immunofluorescence and antibodies directed against the N terminus (Manex1A) or rod domains (Dys1) of dystrophin (Figure 3.1). Immunofluorescence showed no dystrophin staining in all four dystrophic dogs (Figure 3.1). To confirm membrane integrity, an antilaminin antibody was used to stain all muscle biopsies (including the only available muscle sample for the

Figure 3.1. Immunofluorescent analysis of muscle tissue from dystrophic dogs.

Skeletal muscle sections were stained with monoclonal antibodies against Laminin, the N terminus of dystrophin (Manex1a), or the rod domain of dystrophin (Dys1). Sections are shown from a wild type dog (W) as well as the following dystrophic dog lines with previously undefined mutations: Springer Spaniel (S), Labrador Retriever (L), Tibetan Terrier (T), and Cocker Spaniel (C). Sections from a GRMD dog (G) (exon 6 splice site mutation) and the German short hair pointer (R) (deletion of the entire *DMD* gene) were used as disease controls. Scale bar equals 100 μm.


Laminin
Manex1a
Dys1
W
S
L
T
C
G
R

Tibetan terrier, which was accidentally thawed and refrozen during shipping, resulting in damaged tissue) (Figure 3.1). As a positive control, a muscle biopsy was obtained from a wild type dog and stained for dystrophin using the Manex1A and Dys1 antibodies showing normal staining of dystrophin (Figure 3.1). Disease control muscles were obtained from a GRMD dog, known to contain a point mutation in the splice acceptor site of intron 6 (Sharp et al., 1992) resulting in a premature stop codon, as well as a German Short Hair Pointer that was shown to have the entire *DMD* gene deleted (Schatzberg et al., 1999) and both showed no dystrophin staining (Figure 3.1). Immunoblot analysis of proteins isolated from each muscle biopsy showed the presence of full length dystrophin (ca. 400 kDa) in the wild type dog sample, but no dystrophin staining in any of the dystrophic dog samples (Figure 3.2A). Antibodies against myosin heavy chain were used as a loading control during immunoblot analysis (Figure 3.2B)

Mutation analysis of dystrophic dogs

Labrador retriever

To determine the causative mutation in the Labrador retriever, RNA was isolated from the muscle biopsy and cDNA of the *DMD* mRNA was synthesized. Sequencing of the entire *DMD* cDNA revealed a 184 nucleotide (nt) pseudoexon insertion between exon 19 and exon 20 (Figure 3.3A) which results in a premature stop codon immediately downstream of the insertion (p.Gly796X). Using BLAST, the first 19 nt of the insertion were identified as a portion of intron 19 beginning 2,947 nt into the intron with the remainder of the insertion as a Long Interspersed Element (LINE) sequence (Figure 3.3A, Appendix D). A set of intronic primers surrounding the insertion point was used to amplify a 6,838 nt product from genomic template DNA. Internal primers were used to

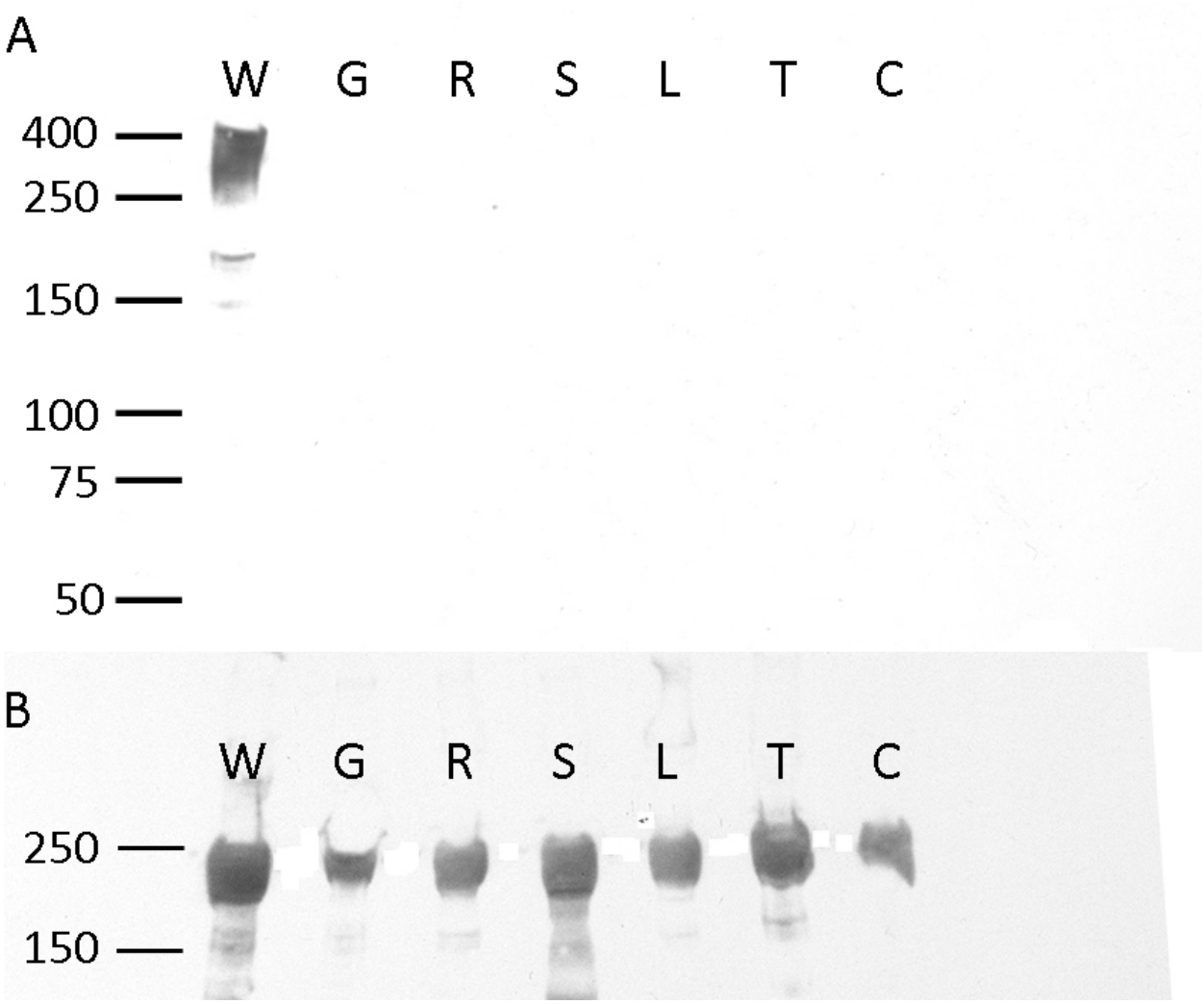


Figure 3.2. Immunoblot analysis of dystrophic dogs.

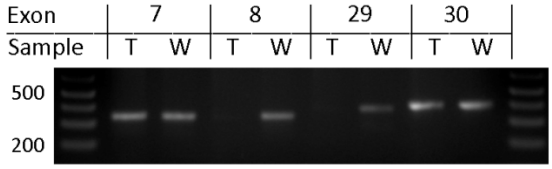
A. Immunoblotting results using antibodies against the rod domain of dystrophin (Dys1) in dystrophic dog lines are shown. Full length dystrophin (ca. 400 kDa) was detected in the protein extract of the wild type muscle (W). No dystrophin was seen for the dystrophic Springer spaniel (S), Labrador retriever (L), Tibetan terrier (T), or Cocker spaniel (C) or the GRMD dog (G) and German short hair pointer (R) disease controls. Size markers are measured in kDa.
B. Immunoblotting results using antibodies against myosin heavy chain are shown for each sample as a loading control.

Figure 3.3. Mutations found in the dystrophic dog lines.

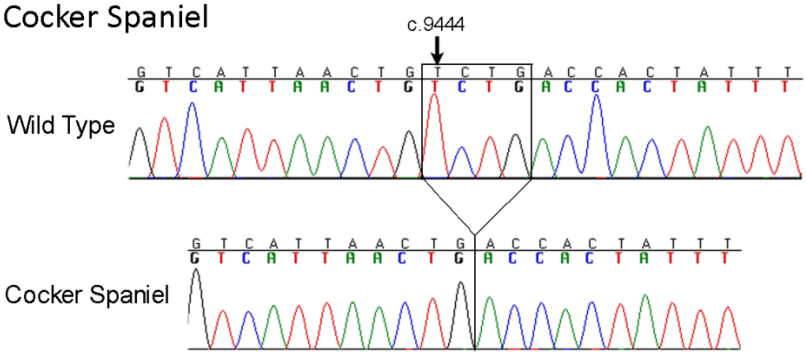
A. LINE-1 insertion in the Labrador retriever. Schematic of the 6,363 nucleotide LINE-1 insertion found in intron 19 of the *DMD* gene and the resulting 184 nt pseudoexon insertion in the *DMD* mRNA.
B. An exon 8 through 29 deletion was found in the Tibetan terrier. Absence of PCR products from exons 8 and 29 of the *DMD* gene in the Tibetan terrier (T), with products of expected size in the wild type dog (W) is shown.
C. Chromatograph of the four nucleotide deletion in exon 65 in the Cocker spaniel is shown. A four nucleotide deletion in exon 65 of the *DMD* gene (lower chromatogram) was found in the Cocker spaniel, compared to the sequencing results from a wild type dog (upper chromatogram).

A. Labrador Retriever

B. Tibetan Terrier



C. Cocker Spaniel



sequence the insertion and surrounding intronic sequence, and we identified the insertion as a 6,363 nt *Canis familiaris* Long Interspersed Element-1 (L1_Cf) transposable element. The LINE is essentially full length; the inserted sequence starts from position 9 of the L1-Cf consensus sequence and is less than 1% divergent from the consensus. An 11 nucleotide sequence from the intron is repeated on both sides of the insertion and matches a target site duplication (TSD) associated with canonical L1 insertions. A "GT" at the 166-167 nt of the L1 insertion serves as a cryptic splice donor site and an "AG" present at position 2945-2946 nt into intron 19 serves as a splice acceptor site for the pseudoexon insertion (Figure 3.3A, Appendix D). The nomenclature of the mutation at the RNA level is thus r.2383_2384ins2383+2949_2383+2967ins165.

Tibetan terrier

Sequencing the entire *DMD* cDNA derived from skeletal muscle mRNA from a dystrophic Tibetan terrier revealed several alternatively spliced products containing deletions of varying numbers of exons found at the 5' end of the gene. The smallest deletion detected among these products encompassed exons 8 through 29; no products were seen that included any exons between exons 8 through 29. Primers were designed to amplify exons 3 through 50 from genomic template DNA obtained from muscle samples of the Tibetan terrier and a wild type control (Appendix D). Products were not obtained for exons 8 through 29 in the Tibetan terrier (Figure 3.3B), whereas all products were amplified successfully from the wild type control. PCR was repeated using an alternate, nonoverlapping set of primers for both exons 8 and 29 in order to exclude the possibility that the absence of amplification was due to a polymorphism at the site of hybridization of one of the original primers (see Materials and methods).

Cocker spaniel

To determine the causative mutation in a dystrophic Cocker spaniel, the entire *DMD* cDNA derived from skeletal muscle mRNA was sequenced and a four nucleotide deletion in exon 65 (c.9444_9447delTCTG) was identified (Figure 3.3C), corresponding to positions c. 9459_9462 in the human *DMD* cDNA. The resulting shift in the reading frame results in a premature stop codon at the site of the deletion (p.Cys3148X).

Springer spaniel

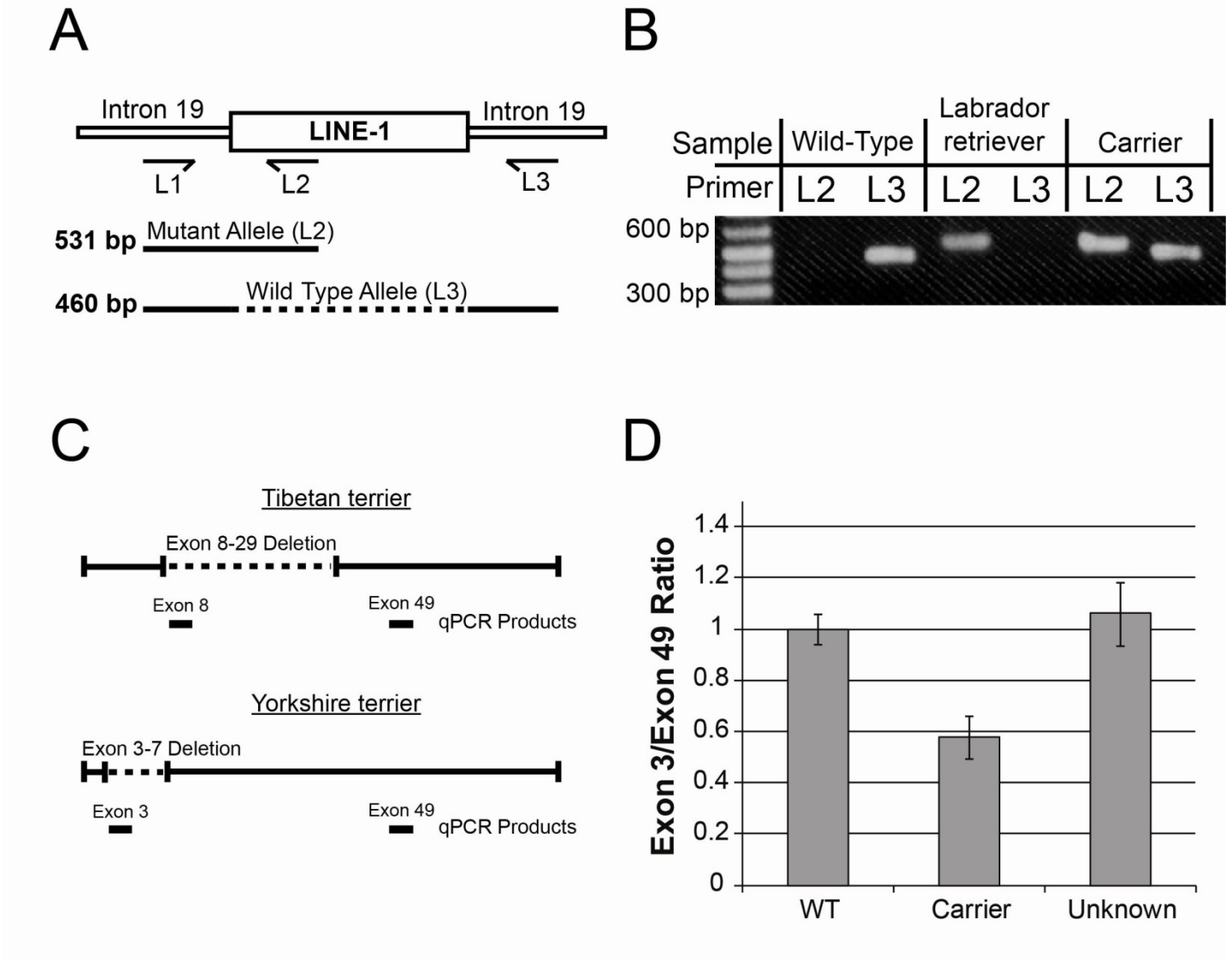
To determine the causative mutation in the dystrophic Springer spaniel, the complete *DMD* cDNA derived from skeletal muscle mRNA was sequenced. However, no mutations were found. Genomic DNA was extracted from muscle tissue and the *DMD* 5' UTR, Dp427m muscle promoter, and 3' UTR were sequenced. No mutations were found in these regions.

Genotyping canine models of DMD

Rapid genotyping assays using extracted genomic DNA were developed for each mutation identified. To assay the presence of the LINE-1 insertion in intron 19 of the *DMD* gene identified in the dystrophic Labrador retrievers, primer pairs were designed to differentiate between the wild type intron 19 or to amplify a portion of the LINE-1 insertion from a colony of dogs (Figure 3.4A) with the wild type allele expected to yield a 460 bp product and the mutant allele expected to yield a 531 bp product in our assay (Figure 3.4A). The assay was successfully used to genotype Labrador retrievers containing the LINE-1 insertion (Figure 3.4B). To detect the presence of the 4 bp deletion in exon 65 found in the Cocker spaniel, the surrounding region was amplified

Figure 3.4. Genotyping dog models of DMD.

A. Schematic of the LINE-1 insertion PCR assay for the Labrador retriever. Primers were designed to amplify either the mutant allele containing the LINE-1 insertion in intron 19 of the *DMD* gene (Primers L1 and L2) or the wild type allele (Primers L1 and L3) from genomic DNA. Primer pair L1/L2 amplifies a 531 bp product if the LINE-1 insertion is present, and primer pair L1/L3 amplifies a 460 bp product with no LINE-1 insertion.
B. Samples genotyped using the LINE-1 insertion PCR assay is shown. A wild type dog sample contains only the wild type allele (L3). The dystrophic male Labrador retriever contains only the mutant allele (L2). A female carrier for the LINE-1 insertion contains both the mutant (L2) and wild type (L3) alleles.
C. Schematic of the qPCR assays used for the deletions found in the Tibetan terrier and Yorkshire terrier. qPCR using genomic DNA was used to calculate the ratio of exon 8 to exon 49 (Tibetan terrier) or exon 3 to exon 49 (Yorkshire terrier) in the *DMD* gene to determine the carrier status of dog samples for the described deletions.
D. A Yorkshire terrier is genotyped. qPCR was used to calculate the ratio of exon 3 to exon 49 in a line of Yorkshire terriers containing an exon 3 through 7 deletion of the *DMD* gene. A female carrier for the deletion contains 57% (+/- 8%) the amount of exon 3 compared to exon 49. A female Yorkshire terrier of unknown status was genotyped and had 106% (+/- 12%) the ratio of exon 3 to exon 49. The unknown sample was determined to be wild type. Levels were normalized to the ratio of exon 3 to exon 49 in a wild type sample (WT).


A
Intron 19
LINE-1
Intron 19
L1
L2
L3
531 bp
Mutant Allele (L2)
460 bp
Wild Type Allele (L3)
B
Sample
Wild-Type
Labrador retriever
Carrier
Primer
L2
L3
L2
L3
L2
L3
600 bp
300 bp
C
Tibetan terrier
Exon 8-29 Deletion
Exon 8
Exon 49
qPCR Products
Yorkshire terrier
Exon 3-7 Deletion
Exon 3
Exon 49
qPCR Products
D
Exon 3/Exon 49 Ratio
1.4
1.2
1
0.8
0.6
0.4
0.2
0
WT
Carrier
Unknown

and sequenced using specific primers (see Materials and methods).

To determine the carrier status of a line of Yorkshire terriers previously known to carry a deletion of exon 3 through exon 7 and the line of Tibetan terriers with a deletion of exon 8 through exon 29 in the *DMD* gene, quantitative real time PCR using SYBR green was performed. Primers were designed to amplify exon 49 of the *DMD* gene as a reference and exon 3 of the *DMD* gene in the Yorkshire terrier or exon 8 in the *DMD* gene in the Tibetan terrier as a test exon (Figure 3.4C). The relative levels of exon 3 or exon 8 compared to exon 49 were used to determine whether the sample was a carrier for the exon 3-7 deletion or for the exon 8-29 deletion (Figure 3.4C). The relative levels of exon 3 or exon 8 to exon 49 were normalized to a wild type sample. For the Yorkshire terrier, a known carrier for the exon 3-7 deletion had 57% the levels of exon 3 compared to exon 49 (Figure 3.4D). An example of a Yorkshire terrier from a dog colony was genotyped and had 106% levels of exon 3 compared to exon 49, which we determined to be wild type (Figure 3.4D).

Discussion

We have identified the *DMD* mutations responsible for three new dog models of Duchenne muscular dystrophy. These mutations represent three different mutation classes found in human dystrophinopathies, including a pseudoexon insertion into mRNA; a small 4 nucleotide frameshifting deletion; and a large out-of-frame deletion spanning 22 exons. These dogs add to the catalog of available models that can be used to study human Duchenne muscular dystrophy.

The L1_Cf insertion we describe in the Labrador retriever appears to be a new insertion belonging to the Long Interspersed Element-1 (LINE-1 or L1) lineage of

transposable elements. L1s are retrotransposons that are transcribed as an RNA intermediate, and then reverse transcribed back into a different position of the genome. Insertions of other repeat elements, such as ALUs and other SINEs, require the L1 reverse transcriptase to initiate mobilization (Chen et al., 2007; Dewannieux et al., 2003). In dogs, L1s comprise more than 14% of the genome and are believed to be more active than human L1s (Wang and Kirkness, 2005). L1 and L1 mediated insertions can impact the gene expression of surrounding genes, and have been shown to cause both canine and human disease (Babushok and Kazazian, 2007). Recently, an L1 insertion in intron 13 of the *DMD* gene was shown to cause muscular dystrophy in a Pembroke Welsh corgi dog (Smith et al., 2011). Other examples of disease causing retrotransposon insertions in dogs include a LINE insertion in the *Factor IX* gene associated with mild form of hemophilia B (Brooks et al., 2003); a short interspersed element (SINE) insertion in the *SILV* gene responsible for merle patterning (Clark et al., 2006); and a SINE insertion in the *PTPLA* associated with centronuclear myopathy in Labradors (Pele et al., 2005). In humans, retrotransposon insertions have been associated with diseases (Babatz and Burns, 2013; Kazazian et al., 1988; Mine et al., 2007), and several insertions have been found within the *DMD* gene (Ferlini et al., 1998; Ferlini and Muntoni, 1998; Holmes et al., 1994; McNaughton et al., 1993; Muntoni et al., 1993; Musova et al., 2006; Narita et al., 1993; Pizzuti et al., 1992), most of which are disease causing. Examples include a chimeric L1 insertion into exon 48 in a patient with disease intermediate in severity between DMD and BMD (Holmes et al., 1994). Insertions of rearranged (Musova et al., 2006) and truncated L1 (Narita et al., 1993) elements into exon 44 lead to exon 44 skipping and to DMD. In contrast, in the Labrador retriever, insertion of the full length

L1 element into intron 19 results in the inclusion of an out-of-frame pseudoexon within the mRNA. This is an increasingly recognized category of human mutation (Gurvich et al., 2008), and the Labrador may prove to be a useful model for preclinical trials directed toward mRNA splicing therapies.

The Tibetan terrier with an out-of-frame deletion spanning 22 exons (exons 8-29) represents one of the only dog models affected by the most common mutation class found in human dystrophinopathy, where deletions of one or more exons account for around 60-65% of mutations (Flanigan et al., 2009). Typically, out-of-frame deletions result in DMD (Monaco et al., 1988), whereas deletions in frame resulting in a modified dystrophin protein with an intact C terminal domain can lead to the less severe BMD, even in the setting of very large deletions (England et al., 1990). Only one patient with an out-of-frame exon 8-29 deletion has been identified and reported to the Leiden Muscular Dystrophy Database (White and den Dunnen, 2006), although the phenotype is not described.

Small out-of-frame deletions similar to the 4 nt deletion in exon 65 found in the Cocker spaniel represent another class of mutations in DMD patients that account for 7.3% of *DMD* mutations in human patients (Flanigan et al., 2009). The Leiden Muscular Dystrophy database includes three 4 nt deletions in exon 65, all of which have been reported as associated with DMD: c.9454_9457del, c.9459_9462del, and c.9471_9474del(Aartsma-Rus et al., 2006) (also reported elsewhere (Gardner et al., 1995)). The canine mutation c.9444_9447delTCTG deletes the same homologous nucleotides as one of these (c.9459_9462del).

The absence of a detectable mutation in the Springer spaniel also has a correlate

in human disease. Sequencing of cDNA derived from muscle mRNA should detect all categories of coding region mutations, but rare patients (0.2% in one large series) (Flanigan et al., 2009) have been reported with absence of dystrophin expression but no *DMD* mutation detectable by mRNA analysis. These include cases where the 5'UTR, including the known promoter, has been sequenced, as was the case in the Springer spaniel. Expression and variation of other genes have been shown to modify the disease phenotype in DMD patients (Flanigan et al., 2012; Heller et al., 2013; Martin et al., 2014; Pegoraro et al., 2011). Perhaps unidentified mutations in other genes necessary for dystrophin expression explain the result in these patients. Alternatively, unidentified mutations could exist in the noncoding regions of the *DMD* gene that alter expression of dystrophin and lead to a DMD phenotype. Mutations in the *DMD* gene that alter splicing signals have been shown to affect disease severity in human patients (Juan-Mateu et al., 2013). Further investigation of the dystrophic Springer spaniel could help determine the cause of this unusual observation as seen in this small subset of patients.

These new models offer new tools for the preclinical development of therapeutic strategies. A current strategy that uses antisense oligonucleotides (AONs) to mediate exon skipping in order to restore the open reading frame and lead to a milder BMD phenotype (Aartsma-Rus and van Ommen, 2007; Gurvich et al., 2008). The Tibetan terrier we describe could be used as a model for AON mediated exon skipping; the open reading frame of the dystrophin mRNA could be restored by skipping either exons 6 and 7 or exons 30-50. Similarly, in the Cocker spaniel, skipping of exons 65 and 66 would restore the open reading frame. These dogs may serve as improved models of nonmutation class specific therapies, as they appear to more faithfully recapitulate human

DMD at the histopathologic level, with more consistent absence of dystrophin expression than is found in the more common GRMD model.

Materials and methods

Immunofluorescence

Frozen muscle biopsies were obtained from Joe Kornegay, Dan Bogen, and Janet Bogen (University of North Carolina). Frozen muscle was cut in 10 micron sections and fixed using 1% paraformaldehyde for 1 minute. Sections were then blocked in 10% fetal calf serum and permeabilized in 0.1% Triton for 30 minutes. Next, the sections were incubated for one hour at room temperature with either an antidystrophin antibody specific to the rod domain (1:10, Dys1, Novocastra, produced in mouse) or to exon 1 (1:10, Manex1a, Clone No. 4C7, a gift of Dr. G.E. Morris, produced in mouse) of dystrophin and washed with phosphate buffer saline (PBS). To evaluate membrane integrity of each sample, all sections were stained with an antilaminin antibody (1:50, Sigma, produced in rabbit). Goat antimouse AlexaFluor 488 (Molecular Probes) and Goat antirabbit AlexaFluor 568 (Molecular Probes) were used as a secondary antibodies for all sections at a 1:500 dilution. Images were obtained on the Olympus IX71 microscope under 20X magnification.

Western blot analyses

Frozen muscle was cut into 10 micron sections and solubilized in 4.6 mM Tris, pH 7.4, 9% SDS, 4% glycerol, 5% B-mercaptoethanol using the TissueLyser (Qiagen) and protein levels were quantified using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). Protein (20 ug per lane) was separated using a Nupage 3-8% Tris-

Acetate precast gel (Invitrogen) and transferred to nitrocellulose. Blots were blocked in Tris-buffered saline pH 7.4 with 0.1% Tween-20 and 5% nonfat milk. Blots were incubated for one hour at room temperature with monoclonal antibodies against dystrophin rod domain (1:100, DYS1, Novocastra) or myosin heavy chain (1:500, MHC H-300, Santa Cruz Biotechnology, Inc.). After washing in Tris buffered saline pH 7.4 with 0.1% Tween-20, secondary antibodies (antimouse IgG at 1:300000 for DYS1 and antirabbit IgG at 1:15000 for antimyosin heavy chain) conjugated to horseradish peroxidase were added for 45 minutes. Blots were washed again and developed by incubating the membrane in Luminol/Enhancer Solution and stable peroxide buffer (Supersignal® West Femto Maximum Sensitivity Substrate kit, Pierce) for 5 minutes and exposing to autoradiography film (Kodak) for one minute.

Genetic analysis of *DMD* mRNA

Frozen muscle was cut into 30-50 sections of 10 microns each and total RNA was extracted using Trizol (Invitrogen, La Jolla, CA) according to the manufacturer's recommended protocol. RNA levels were quantified using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). Complementary DNA (cDNA) was synthesized using random hexamers and SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's recommended protocol. Primer sets were designed to amplify the coding region of the *DMD* cDNA in 10 amplicons (i.e., sections) (Table 3.1). Due to the relatively low concentration of *DMD* mRNA in RNA extracts, each amplicon was amplified in two PCR steps using external and internal primer sets using Expand High Fidelity Taq polymerase (Roche) according to the manufacturer's recommended protocol (Table 3.1). Each section was sequenced using an internal set of sequencing primers

Table 3.1

Primer sets used to amplify and sequence the canine *DMD* mRNA.

Amplicon	Primer Set	Forward Primer	Reverse Primer
1	External Primer Set	GGCTGCTGAAGTTGGTTGAT	TGGTGACATAAGCAGCCTGT
	Internal Primer Set	TTGGAGCACTTAAGTTGGGAG	GTGTGTAGGCATAGCTCTTGAA
	Sequencing Primers	CAACTCGTAATTATCCACAGGTT	CTGTGGATGAGAGCGTTCAA
2	External Primer Set	TCATCAAATGCACTATTCTCAACA	AGCTGATCTGCTGGCATCTT
	Internal Primer Set	CACGGTCAGTCTAGCACAGG	TCTCGCTCTATGGCATTGACT
	Sequencing Primers	TGGTGGTGGTAGTCGATGAA	TTGTTCTTCCAAAGCAGCAG
3	External Primer Set	AGAAGCTGTGTTGCAGAGTCC	TGGGCCTCTTCTTTAGCTCTC
	Internal Primer Set	TGCAATCTTTCGGAAGGAAG	TTCTTCAGCTTGTGTCATCCA
	Sequencing Primers	GAGAAAGGACAAGGGCCAAT	TGATTTGTAAAGGCCACAAAGTC
4	External Primer Set	TTTTCCTGAAGGAGGAATGG	TGCTTCTTCTGTACAATCTGACG
	Internal Primer Set	AAAGACAGCTGAAACAGTGCAG	TCAGACTTCACTTCACTCAGACTT
	Sequencing Primers	CTTGGAGAAGGCAAACAAGTG	TTCAAGCGAATCAAGCACCT
5	External Primer Set	GCCTGCATTGGAAACAAAGA	TGTCATCCAAGCATTTCAGG
	Internal Primer Set	CAGTCACAGCTAAATCATTGTGTG	ATCATCAGCCTGCCTCTTGT
	Sequencing Primers	CCCGAGCAGAAGAGTGGTTA	TGATGTAATCCACATTCTGGTCA
6	External Primer Set	GAATTGTTGCAAAGAGGAGACA	AGCTGTTTGCAGACCTCCTG
	Internal Primer Set	CAGATGAGAGAAAGCGAGAGG	TTTCTTCCCCAGTTGCATTC
	Sequencing Primers	AGAGTCAGTGGTGGCGATG	CAAAGGTCGGGAGCATTAAG
7	External Primer Set	TGGCTAACAGAAGCTGAACAG	CCAGAGCAGGTACCTCCAAC
	Internal Primer Set	CCTGAGAATTGGGAACATGC	CCTTAGTAACCACAGGTTGTGTCA
	Sequencing Primers	GCTCCCATAAGCCCAGAAG	AGGAGATAACCACAGCAGCAG
8	External Primer Set	TTCAAGAGCTGAGGGCAAAG	TCTCTGGGCTCCTGGTAGAG
	Internal Primer Set	GACCTAGCTCCTGGACTGACC	GCTGCTCTGTCAGAAATATTCG
	Sequencing Primers	CGCCAGTGGCAGATAAATGT	CCCAAGAGGCATTGATGTTC
9	External Primer Set	GGAAGCCAGTTCTGACCAGT	CTTGCAGATGTTGCACTTGG
	Internal Primer Set	GAAGCGTCTGCACCTTTCTC	GCGCATCCAGTCTAGGAAGA
	Sequencing Primers	CCAGCACTTCCTTTCCACTT	AGGGCACTTTGTTTGGTGAG
10	External Primer Set	TCGACAGGATTTTGTGACCA	GCGGGAATCAGGAGTTGTAA
	Internal Primer Set	CCCAAGACAGTTGGGTGAAG	CAAATCATCTGCCAAGTGGA
	Sequencing Primers	ATCCAGCATTACTGGCGAAG	GATCTGGGCAGGACTACGAG

Note: The primers used to amplify and sequence the *DMD* mRNA in dogs are shown. Due to the relatively low concentration of *DMD* mRNA, each amplicon was amplified twice using an external and internal primer set. Internal forward and reverse sequencing primers were also designed for each amplicon.

(Table 3.1) using BigDye v3.1 sequencing reagent (Applied Biosystems) on the ABI 3700 DNA analyzer (Applied Biosystems, Foster City, CA) prepared with POP-5 capillary gel matrix. The entire coding region of the *DMD* cDNA was sequenced for each dog and sequence files were analyzed using Sequencher (Gene Codes, Ann Arbor, MI) and compared against the wild type canine dystrophin cDNA sequence (NM 001003343).

Genetic analysis of genomic DNA

Where appropriate, genomic DNA was analyzed to verify mutations identified from mRNA. DNA was isolated from 30 ten micron sections of frozen muscle using the Puregene Genomic DNA Purification Kit (Qiagen) according to the manufacturer's recommended protocol. DNA levels were quantified using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). PCR amplification was conducted using Taq DNA Polymerase (Roche) according to the manufacturer's protocol for short products (100-500 nts). PCR products were sequenced using internal sequencing primers and the BigDye v3.1 sequencing reagent (Applied Biosystems) on the ABI 3700 DNA analyzer (Applied Biosystems, Foster City, CA) prepared with POP-5 capillary gel matrix. Sequence files were analyzed using Sequencher (Gene Codes, Ann Arbor, MI).
The entire LINE-1 insertion in the Labrador retriever was amplified using the flanking intronic primers TGAAAGTAAGAGCTGAGTCATGG and TCGCCAAAAGTGAATTGAAA with the Expand Long Range DNA polymerase (Roche) according to the manufacturer's recommended protocol. The entire LINE-1 insertion was sequenced using internal sequencing primers as described above.

To determine the boundaries of the deletion in the Tibetan terrier, primers were

designed to amplify exons 3 through 32 and exons 44 through 50 of the *DMD* gene from genomic DNA (Appendix D). An alternate set of primers for exons 8 and 29 were made to verify the deletion extended to exons 8 and 29 (Appendix D). PCR amplification was done using Taq DNA Polymerase (Roche) according to the manufacturer's protocol and electrophoresed on a 1% agarose gel stained with ethidium bromide.

The muscle promoter region, *DMD* 5' UTR and *DMD* 3' UTR was sequenced from genomic DNA obtained from the Springer spaniel. The 5' region of the muscle isoform of dystrophin including the promoter and 5' UTR (from 1,285 nucleotides upstream of the start codon) was amplified using the primers TTCTGTGCCAAGCAATTGAA and GGGCATGAACTCTTGTGGAT using Taq DNA Polymerase (Roche) according the manufacturer's protocol and sequenced using internal primers as described above. The *DMD* 3' UTR was amplified in four sections using the following primers: Section 1 – primers TGGTACAACGGTGTCTTCTCC and CAGCCAGTTCCAATCAGTCA; Section 2 – primers TATGGAACGCTTTTGGGTTG and TTACCAGGAGCACACCATGA; Section 3 – primers TCCCAGACAGTGAGGAGGAC and TGCACGCTATTTACCTCTGC; and Section 4 – primers ACATCCAACATGGCTTCTCA and TGCGTGCTTTATTGAGATACAC. Each section of the 3' UTR was sequenced using internal primers as described above.

Genotyping mutations in dogs

Genomic DNA was isolated from either 30 ten micron sections of frozen muscle using the Puregene Genomic DNA Purification Kit (Qiagen) according to the manufacturer's recommended protocol or from whole blood using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's recommended protocol for all

genotyping assays described. DNA levels were quantified using the Nanodrop ND-1000 Spectrophotometer (Thermo Scientific). To determine the carrier status of the LINE-1 insertion in intron 19 of the *DMD* gene found in the Labrador retriever, a PCR assay was used to differentially amplify the wild type or mutant allele. The primers L1 (TGAAAGTAAGAGCTGAGTCATGG) and L2 (GAAGCGCGAACTCTTCTCAC) were designed to amplify a 531 bp product if the LINE-1 insertion was present, and the primers L1 (TGAAAGTAAGAGCTGAGTCATGG) and L3 (TCGCCAAAAGTGAATTGAAA) were designed to amplify a 460 bp product if the LINE-1 insertion was absent. PCR amplification was conducted using Taq DNA Polymerase (Roche) according to the manufacturer's recommended protocol, and products were electrophoresed on a 1% agarose gel.

To determine the presence of the 4 bp deletion in exon 65 of the *DMD* gene found in the Cocker spaniel, the primers TGGATCTCTTGAGCCTATCG and ATTCAGCAGCCAATTGAGAC were used to amplify the region of interest from genomic DNA. PCR amplification was conducted using Taq DNA Polymerase (Roche) according to the manufacturer's recommended protocol, and PCR products were sequenced as described above.

Quantitative PCR (qPCR) was used to detect the carrier status of the deletion of exons 3 through 7 of the *DMD* gene found in the Yorkshire terrier or the deletion of exons 8 through 29 found in the Tibetan terrier by comparing the relative levels of a test *DMD* exon deleted in each mutation compared to a reference exon in the same gene. The primers TGCCCTTCCTAGAATCAACA and GTCTCCCATCCTGTAGGTCA were used to amplify *DMD* exon 3 in the Yorkshire terrier and primers

TTGTTCATTCGCCTTTCAGA and TGTGTATGATGGCTAAAAATGC were used to amplify *DMD* exon 8 in the Tibetan terrier. The reference *DMD* exon 49 was amplified using the primers CCCAGGAAATTGAAGTAGCA and AGCACACCAGGACAGAATTG in both dogs. qPCR reaction mixtures contained a final concentration of 1x Platinum SYBR Green Mix (Invitrogen), 20 ng of genomic DNA, and 0.5 μM of forward and reverse primers and done in replicates of four. After an initial denaturation of 95°C for 2 minutes, amplification was performed using the following cycle conditions: 40 cycles of 95°C for 15 seconds, 58°C for 30 seconds, and 72°C for 30 seconds. Fluorescence intensity was monitored once per cycle after the elongation phase using the 7900HT (Applied Biosystems). A melting curve was produced after each run starting from 60°C to 95°C at a ramp rate of 2% on the 7900 HT (Applied Biosystems). A standard curve was produced for each primer pair by diluting the control DNA to concentrations of 40 ng, 20 ng, 10 ng, and 5 ng of DNA and running four duplicates for each concentration. Results were analyzed using SDS 2.3 software (Applied Biosystems) and calculated using the comparative C_T ($\Delta\Delta C_T$) method.

Acknowledgements

I would like to thank Dr. Kevin Flanigan (Nationwide Children's Hospital, Columbus, Ohio) for his supervision and guidance on this work. I would like to thank Joe Kornegay (University of North Carolina), Dan Bogen (University of North Carolina), and Janet Bogen (University of North Carolina) for providing frozen muscle tissue from dystrophic dogs. I would also like to thank Laura Taylor and Baijayanta "Jay" Maiti for technical support.

References

Aartsma-Rus, A., Van Deutekom, J.C., Fokkema, I.F., Van Ommen, G.J., and Den Dunnen, J.T. (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve *34*, 135-144.

Aartsma-Rus, A., and van Ommen, G.J. (2007). Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. RNA *13*, 1609-1624.

Babatz, T.D., and Burns, K.H. (2013). Functional impact of the human mobilome. Curr Opin Genet Dev *23*, 264-270.

Babushok, D.V., and Kazazian, H.H., Jr. (2007). Progress in understanding the biology of the human mutagen LINE-1. Hum Mutat *28*, 527-539.

Baltzer, W.I., Calise, D.V., Levine, J.M., Shelton, G.D., Edwards, J.F., and Steiner, J.M. (2007). Dystrophin-deficient muscular dystrophy in a Weimaraner. J Am Anim Hosp Assoc *43*, 227-232.

Baroncelli, A.B., Abellonio, F., Pagano, T.B., Esposito, I., Peirone, B., Papparella, S., and Paciello, O. (2013). Muscular Dystrophy in a Dog Resembling Human Becker Muscular Dystrophy. J Comp Pathol.

Barthelemy, I., Thibaud, J., Carelle, N., Tiret, L., Vulin, A., Drouard, C., Garcia, L., Kaplan, J., Leturcq, F., and Blot, S. (2007). Clinical description and molecular analysis of a promising model of DMD, the LRMD dog. Neuromuscular Disorders *17*, 776-776.

Barton-Davis, E.R., Cordier, L., Shoturma, D.I., Leland, S.E., and Sweeney, H.L. (1999). Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of mdx mice. J Clin Invest *104*, 375-381.

Bassett, D., and Currie, P.D. (2004). Identification of a zebrafish model of muscular dystrophy. Clin Exp Pharmacol Physiol *31*, 537-540.

Baumeister, R., and Ge, L. (2002). The worm in us - Caenorhabditis elegans as a model of human disease. Trends Biotechnol *20*, 147-148.

Bergman, R.L., Inzana, K.D., Monroe, W.E., Shell, L.G., Liu, L.A., Engvall, E., and Shelton, G.D. (2002). Dystrophin-deficient muscular dystrophy in a Labrador retriever. J Am Anim Hosp Assoc *38*, 255-261.

Bessou, C., Giugia, J.B., Franks, C.J., Holden-Dye, L., and Segalat, L. (1998). Mutations in the Caenorhabditis elegans dystrophin-like gene dys-1 lead to hyperactivity and suggest a link with cholinergic transmission. Neurogenetics *2*, 61-72.

Blake, D.J., Weir, A., Newey, S.E., and Davies, K.E. (2002). Function and genetics of dystrophin and dystrophin-related proteins in muscle. Physiol Rev *82*, 291-329.

Blot, S., Carelle, N., Deburgrave, N., Escriou, C., Chetboul, V., Kaplan, J.C., and Leturcq, F. (2002). LRMD: A new canine model of dystrophinopathy in a Labrador retriever strain. J Neurol Sci *199*, S9-S9.

Brooks, M.B., Gu, W., Barnas, J.L., Ray, J., and Ray, K. (2003). A Line 1 insertion in the Factor IX gene segregates with mild hemophilia B in dogs. Mamm Genome *14*, 788-795.

Bulfield, G., Siller, W.G., Wight, P.A., and Moore, K.J. (1984). X chromosome-linked muscular dystrophy (mdx) in the mouse. Proc Natl Acad Sci U S A *81*, 1189-1192.

Carpenter, J.L., Hoffman, E.P., Romanul, F.C., Kunkel, L.M., Rosales, R.K., Ma, N.S., Dasbach, J.J., Rae, J.F., Moore, F.M., McAfee, M.B.*, et al.* (1989). Feline muscular dystrophy with dystrophin deficiency. Am J Pathol *135*, 909-919.

Chambers, S.P., Dodd, A., Overall, R., Sirey, T., Lam, L.T., Morris, G.E., and Love, D.R. (2001). Dystrophin in adult zebrafish muscle. Biochem Biophys Res Commun *286*, 478-483.

Chen, J.M., Ferec, C., and Cooper, D.N. (2007). Mechanism of Alu integration into the human genome. Genomic Med *1*, 9-17.

Cirak, S., Arechavala-Gomeza, V., Guglieri, M., Feng, L., Torelli, S., Anthony, K., Abbs, S., Garralda, M.E., Bourke, J., Wells, D.J.*, et al.* (2011). Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. Lancet *378*, 595-605.

Clark, L.A., Wahl, J.M., Rees, C.A., and Murphy, K.E. (2006). Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. Proc Natl Acad Sci U S A *103*, 1376-1381.

Cohn, R.D., and Campbell, K.P. (2000). Molecular basis of muscular dystrophies. Muscle Nerve *23*, 1456-1471.

Collins, C.A., and Morgan, J.E. (2003). Duchenne's muscular dystrophy: animal models used to investigate pathogenesis and develop therapeutic strategies. Int J Exp Pathol *84*, 165-172.

Cooper, B.J., Winand, N.J., Stedman, H., Valentine, B.A., Hoffman, E.P., Kunkel, L.M., Scott, M.O., Fischbeck, K.H., Kornegay, J.N., Avery, R.J.*, et al.* (1988). The homologue of the Duchenne locus is defective in X-linked muscular dystrophy of dogs. Nature *334*, 154-156.

Dangain, J., and Vrbova, G. (1984). Muscle development in mdx mutant mice. Muscle Nerve *7*, 700-704.

Deconinck, N., and Dan, B. (2007). Pathophysiology of duchenne muscular dystrophy: current hypotheses. Pediatr Neurol *36*, 1-7.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat Genet *35*, 41-48.

England, S.B., Nicholson, L.V., Johnson, M.A., Forrest, S.M., Love, D.R., Zubrzycka-Gaarn, E.E., Bulman, D.E., Harris, J.B., and Davies, K.E. (1990). Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. Nature *343*, 180-182.

Fabb, S.A., Wells, D.J., Serpente, P., and Dickson, G. (2002). Adeno-associated virus vector gene transfer and sarcolemmal expression of a 144 kDa micro-dystrophin effectively restores the dystrophin-associated protein complex and inhibits myofibre degeneration in nude/mdx mice. Hum Mol Genet *11*, 733-741.

Fairclough, R.J., Wood, M.J., and Davies, K.E. (2013). Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches. Nat Rev Genet *14*, 373-378.

Ferlini, A., Galie, N., Merlini, L., Sewry, C., Branzi, A., and Muntoni, F. (1998). A novel Alu-like element rearranged in the dystrophin gene causes a splicing mutation in a family with X-linked dilated cardiomyopathy. Am J Hum Genet *63*, 436-446.

Ferlini, A., and Muntoni, F. (1998). The 5' region of intron 11 of the dystrophin gene contains target sequences for mobile elements and three overlapping ORFs. Biochem Biophys Res Commun *242*, 401-406.

Finkel, R.S. (2010). Read-through strategies for suppression of nonsense mutations in Duchenne/ Becker muscular dystrophy: aminoglycosides and ataluren (PTC124). J Child Neurol *25*, 1158-1164.

Flanigan, K.M., Dunn, D.M., von Niederhausern, A., Soltanzadeh, P., Gappmaier, E., Howard, M.T., Sampson, J.B., Mendell, J.R., Wall, C., King, W.M.*, et al.* (2009). Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. Hum Mutat *30*, 1657-1666.

Flanigan, K.M., Ceco, E., Lamar, K.M., Kaminoh, Y., Dunn, D.M., Mendell, J.R., King, W.M., Pestronk, A., Florence, J.M., Mathews, K.D.*, et al.* (2013). LTBP4 genotype predicts age of ambulatory loss in Duchenne muscular dystrophy. Ann Neurol 73, 481-488.

Foster, H., Sharp, P.S., Athanasopoulos, T., Trollet, C., Graham, I.R., Foster, K., Wells, D.J., and Dickson, G. (2008). Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in

dystrophic mdx mice following AAV2/8 gene transfer. Mol Ther *16*, 1825-1832.

Foster, H., Popplewell, L., and Dickson, G. (2012). Genetic therapeutic approaches for Duchenne muscular dystrophy. Hum Gene Ther *23*, 676-687.

Gardner, R.J., Bobrow, M., and Roberts, R.G. (1995). The identification of point mutations in Duchenne muscular dystrophy patients by using reverse-transcription PCR and the protein truncation test. Am J Hum Genet *57*, 311-320.

Goemans, N.M., Tulinius, M., van den Akker, J.T., Burm, B.E., Ekhart, P.F., Heuvelmans, N., Holling, T., Janson, A.A., Platenburg, G.J., Sipkens, J.A.*, et al.* (2011). Systemic administration of PRO051 in Duchenne's muscular dystrophy. N Engl J Med *364*, 1513-1522.

Gurvich, O.L., Tuohy, T.M., Howard, M.T., Finkel, R.S., Medne, L., Anderson, C.B., Weiss, R.B., Wilton, S.D., and Flanigan, K.M. (2008). DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy. Ann Neurol *63*, 81-89.

Guyon, J.R., Mosley, A.N., Zhou, Y., O'Brien, K.F., Sheng, X., Chiang, K., Davidson, A.J., Volinski, J.M., Zon, L.I., and Kunkel, L.M. (2003). The dystrophin associated protein complex in zebrafish. Hum Mol Genet *12*, 601-615.

Heller, K.N., Montgomery, C.L., Janssen, P.M., Clark, K.R., Mendell, J.R., and Rodino-Klapac, L.R. (2013). AAV-mediated overexpression of human alpha7 integrin leads to histological and functional improvement in dystrophic mice. Mol Ther *21*, 520-525.

Hoffman, E.P., Brown, R.H., Jr., and Kunkel, L.M. (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell *51*, 919-928.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat Genet *7*, 143-148.

Howard, M.T., Anderson, C.B., Fass, U., Khatri, S., Gesteland, R.F., Atkins, J.F., and Flanigan, K.M. (2004). Readthrough of dystrophin stop codon mutations induced by aminoglycosides. Ann Neurol *55*, 422-426.

Jones, B.R., Brennan, S., Mooney, C.T., Callanan, J.J., McAllister, H., Guo, L.T., Martin, P.T., Engvall, E., and Shelton, G.D. (2004). Muscular dystrophy with truncated dystrophin in a family of Japanese Spitz dogs. J Neurol Sci *217*, 143-149.

Juan-Mateu, J., Gonzalez-Quereda, L., Rodriguez, M.J., Verdura, E., Lazaro, K., Jou, C., Nascimento, A., Jimenez-Mallebrera, C., Colomer, J., Monges, S.*, et al.* (2013). Interplay between DMD point mutations and splicing signals in Dystrophinopathy phenotypes. PLoS One *8*, e59916.

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature *332*, 164-166.

Kinali, M., Arechavala-Gomeza, V., Feng, L., Cirak, S., Hunt, D., Adkin, C., Guglieri, M., Ashton, E., Abbs, S., Nihoyannopoulos, P.*, et al.* (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. Lancet Neurol *8*, 918-928.

Klarenbeek, S., Gerritzen-Bruning, M.J., Rozemuller, A.J., and van der Lugt, J.J. (2007). Canine X-linked muscular dystrophy in a family of Grand Basset Griffon Vendeen dogs. J Comp Pathol *137*, 249-252.

Koo, T., Malerba, A., Athanasopoulos, T., Trollet, C., Boldrin, L., Ferry, A., Popplewell, L., Foster, H., Foster, K., and Dickson, G. (2011a). Delivery of AAV2/9-microdystrophin genes incorporating helix 1 of the coiled-coil motif in the C-terminal domain of dystrophin improves muscle pathology and restores the level of alpha1-syntrophin and alpha-dystrobrevin in skeletal muscles of mdx mice. Hum Gene Ther *22*, 1379-1388.

Koo, T., Okada, T., Athanasopoulos, T., Foster, H., Takeda, S., and Dickson, G. (2011b). Long-term functional adeno-associated virus-microdystrophin expression in the dystrophic CXMDj dog. J Gene Med *13*, 497-506.

Kornegay, J.N., Tuler, S.M., Miller, D.M., and Levesque, D.C. (1988). Muscular dystrophy in a litter of golden retriever dogs. Muscle Nerve *11*, 1056-1064.

Lu, Q.L., Rabinowitz, A., Chen, Y.C., Yokota, T., Yin, H., Alter, J., Jadoon, A., Bou-Gharios, G., and Partridge, T. (2005). Systemic delivery of antisense oligoribonucleotide restores dystrophin expression in body-wide skeletal muscles. Proc Natl Acad Sci U S A *102*, 198-203.

Malik, V., Rodino-Klapac, L.R., Viollet, L., Wall, C., King, W., Al-Dahhak, R., Lewis, S., Shilling, C.J., Kota, J., Serrano-Munuera, C.*, et al.* (2010). Gentamicin-induced readthrough of stop codons in Duchenne muscular dystrophy. Ann Neurol *67*, 771-780.

Martin, P.T., Golden, B., Okerblom, J., Camboni, M., Chandrasekharan, K., Xu, R., Varki, A., Flanigan, K.M., and Kornegay, J.N. (2014). A Comparative Study of N-glycolylneuraminic Acid (Neu5Gc) and Cytotoxic T Cell (CT) Carbohydrate Expression in Normal and Dystrophin-Deficient Dog and Human Skeletal Muscle. PLoS One *9*, e88226.

McNaughton, J.C., Broom, J.E., Hill, D.F., Jones, W.A., Marshall, C.J., Renwick, N.M., Stockwell, P.A., and Petersen, G.B. (1993). A cluster of transposon-like repetitive sequences in intron 7 of the human dystrophin gene. J Mol Biol *232*, 314-321.

Mendell, J.R., Rodino-Klapac, L.R., Sahenk, Z., Roush, K., Bird, L., Lowes, L.P., Alfano, L., Gomez, A.M., Lewis, S., Kota, J.*, et al.* (2013). Eteplirsen for the treatment of Duchenne muscular dystrophy. Ann Neurol *74*, 637-647.

Mine, M., Chen, J.M., Brivet, M., Desguerre, I., Marchant, D., de Lonlay, P., Bernard, A., Ferec, C., Abitbol, M., Ricquier, D.*, et al.* (2007). A large genomic deletion in the PDHX gene caused by the retrotranspositional insertion of a full-length LINE-1 element. Hum Mutat *28*, 137-142.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics *2*, 90-95.

Muir, L.A., and Chamberlain, J.S. (2009). Emerging strategies for cell and gene therapy of the muscular dystrophies. Expert Rev Mol Med *11*, e18.

Muntoni, F., Cau, M., Congiu, R., Congia, M., Cao, A., and Melis, M.A. (1993). Identification of a novel T-insertion polymorphism at the DMD locus. Hum Genet *92*, 103.

Muntoni, F., Torelli, S., and Ferlini, A. (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol *2*, 731-740.

Musova, Z., Hedvicakova, P., Mohrmann, M., Tesarova, M., Krepelova, A., Zeman, J., and Sedlacek, Z. (2006). A novel insertion of a rearranged L1 element in exon 44 of the dystrophin gene: further evidence for possible bias in retroposon integration. Biochem Biophys Res Commun *347*, 145-149.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Minami, R., Nakamura, H., and Matsuo, M. (1993). Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. J Clin Invest *91*, 1862-1867.

Pegoraro, E., Hoffman, E.P., Piva, L., Gavassini, B.F., Cagnin, S., Ermani, M., Bello, L., Soraru, G., Pacchioni, B., Bonifati, M.D.*, et al.* (2011). SPP1 genotype is a determinant of disease severity in Duchenne muscular dystrophy. Neurology *76*, 219-226.

Pele, M., Tiret, L., Kessler, J.L., Blot, S., and Panthier, J.J. (2005). SINE exonic insertion in the PTPLA gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. Hum Mol Genet *14*, 1417-1427.

Pichavant, C., Aartsma-Rus, A., Clemens, P.R., Davies, K.E., Dickson, G., Takeda, S., Wilton, S.D., Wolff, J.A., Wooddell, C.I., Xiao, X.*, et al.* (2011). Current status of pharmaceutical and genetic therapeutic approaches to treat DMD. Mol Ther *19*, 830-840.

Pizzuti, A., Pieretti, M., Fenwick, R.G., Gibbs, R.A., and Caskey, C.T. (1992). A

transposon-like element in the deletion-prone region of the dystrophin gene. Genomics *13*, 594-600.

Rodino-Klapac, L.R., Montgomery, C.L., Bremer, W.G., Shontz, K.M., Malik, V., Davis, N., Sprinkle, S., Campbell, K.J., Sahenk, Z., Clark, K.R.*, et al.* (2010). Persistent expression of FLAG-tagged micro dystrophin in nonhuman primates following intramuscular and vascular delivery. Mol Ther *18*, 109-117.

Schatzberg, S.J., Olby, N.J., Breen, M., Anderson, L.V., Langford, C.F., Dickens, H.F., Wilton, S.D., Zeiss, C.J., Binns, M.M., Kornegay, J.N.*, et al.* (1999). Molecular analysis of a spontaneous dystrophin 'knockout' dog. Neuromuscul Disord *9*, 289-295.

Sharp, N.J., Kornegay, J.N., Van Camp, S.D., Herbstreith, M.H., Secore, S.L., Kettle, S., Hung, W.Y., Constantinou, C.D., Dykstra, M.J., Roses, A.D.*, et al.* (1992). An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. Genomics *13*, 115-121.

Sicinski, P., Geng, Y., Ryder-Cook, A.S., Barnard, E.A., Darlison, M.G., and Barnard, P.J. (1989). The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science *244*, 1578-1580.

Smith, B.F., Yue, Y., Woods, P.R., Kornegay, J.N., Shin, J.H., Williams, R.R., and Duan, D. (2011). An intronic LINE-1 element insertion in the dystrophin gene aborts dystrophin expression and results in Duchenne-like muscular dystrophy in the corgi breed. Lab Invest *91*, 216-231.

Valentine, B.A., Cooper, B.J., de Lahunta, A., O'Quinn, R., and Blue, J.T. (1988). Canine X-linked muscular dystrophy. An animal model of Duchenne muscular dystrophy: clinical studies. J Neurol Sci *88*, 69-81.

van Deutekom, J.C., Janson, A.A., Ginjaar, I.B., Frankhuizen, W.S., Aartsma-Rus, A., Bremmer-Bout, M., den Dunnen, J.T., Koop, K., van der Kooi, A.J., Goemans, N.M.*, et al.* (2007). Local dystrophin restoration with antisense oligonucleotide PRO051. N Engl J Med *357*, 2677-2686.

Walmsley, G.L., Arechavala-Gomeza, V., Fernandez-Fuente, M., Burke, M.M., Nagel, N., Holder, A., Stanley, R., Chandler, K., Marks, S.L., Muntoni, F.*, et al.* (2010). A duchenne muscular dystrophy gene hot spot mutation in dystrophin-deficient cavalier king charles spaniels is amenable to exon 51 skipping. PLoS One *5*, e8647.

Wang, B., Li, J., and Xiao, X. (2000). Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in mdx mouse model. Proc Natl Acad Sci U S A *97*, 13714-13719.

Wang, W., and Kirkness, E.F. (2005). Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res *15*, 1798-1808.

Wang, Z., Kuhr, C.S., Allen, J.M., Blankinship, M., Gregorevic, P., Chamberlain, J.S., Tapscott, S.J., and Storb, R. (2007). Sustained AAV-mediated dystrophin expression in a canine model of Duchenne muscular dystrophy with a brief course of immunosuppression. Mol Ther *15*, 1160-1166.

Wang, Z., Storb, R., Halbert, C.L., Banks, G.B., Butts, T.M., Finn, E.E., Allen, J.M., Miller, A.D., Chamberlain, J.S., and Tapscott, S.J. (2012). Successful regional delivery and long-term expression of a dystrophin gene in canine muscular dystrophy: a preclinical model for human therapies. Mol Ther *20*, 1501-1507.

Wetterman, C.A., Harkin, K.R., Cash, W.C., Nietfield, J.C., and Shelton, G.D. (2000). Hypertrophic muscular dystrophy in a young dog. J Am Vet Med Assoc *216*, 878-881, 864.

White, S.J., and den Dunnen, J.T. (2006). Copy number variation in the genome; the human DMD gene as an example. Cytogenet Genome Res *115*, 240-246.

Wieczorek, L.A., Garosi, L.S., and Shelton, G.D. (2006). Dystrophin-deficient muscular dystrophy in an old English sheepdog. Vet Rec *158*, 270-273.

Willmann, R., Possekel, S., Dubach-Powell, J., Meier, T., and Ruegg, M.A. (2009). Mammalian animal models for Duchenne muscular dystrophy. Neuromuscul Disord *19*, 241-249.

APPENDIX A

MRNA STABILITY OF *DMD* 3' UTR CONSTRUCTS

Introduction

3' UTRs have been shown to regulate gene expression posttranscriptionally by altering mRNA stability, localizing mRNA, or directly affecting translation (Andreassi and Riccio, 2009; Gramolini et al., 2001; Matoulkova et al., 2012; Mazumder et al., 2003). We have shown that the highly conserved elements (HCEs) spanning the Lemaire A and Lemaire D regions of *DMD* 3' UTR regulate translation and mRNA abundance in differentiated C2C12 cells (Chapter 1). It has been shown previously that mRNA stability of the *DMD* transcript has a larger impact on dystrophin protein levels than transcription rate in BMD patients (Spitali et al., 2013) and could impact the effectiveness of developing therapies. Here, we use the transcription inhibitor Actinomycin D to directly measure changes in mRNA stability and show that the HCEs spanning the Lemaire A and D regions increases mRNA stability in C2C12 myotubes.

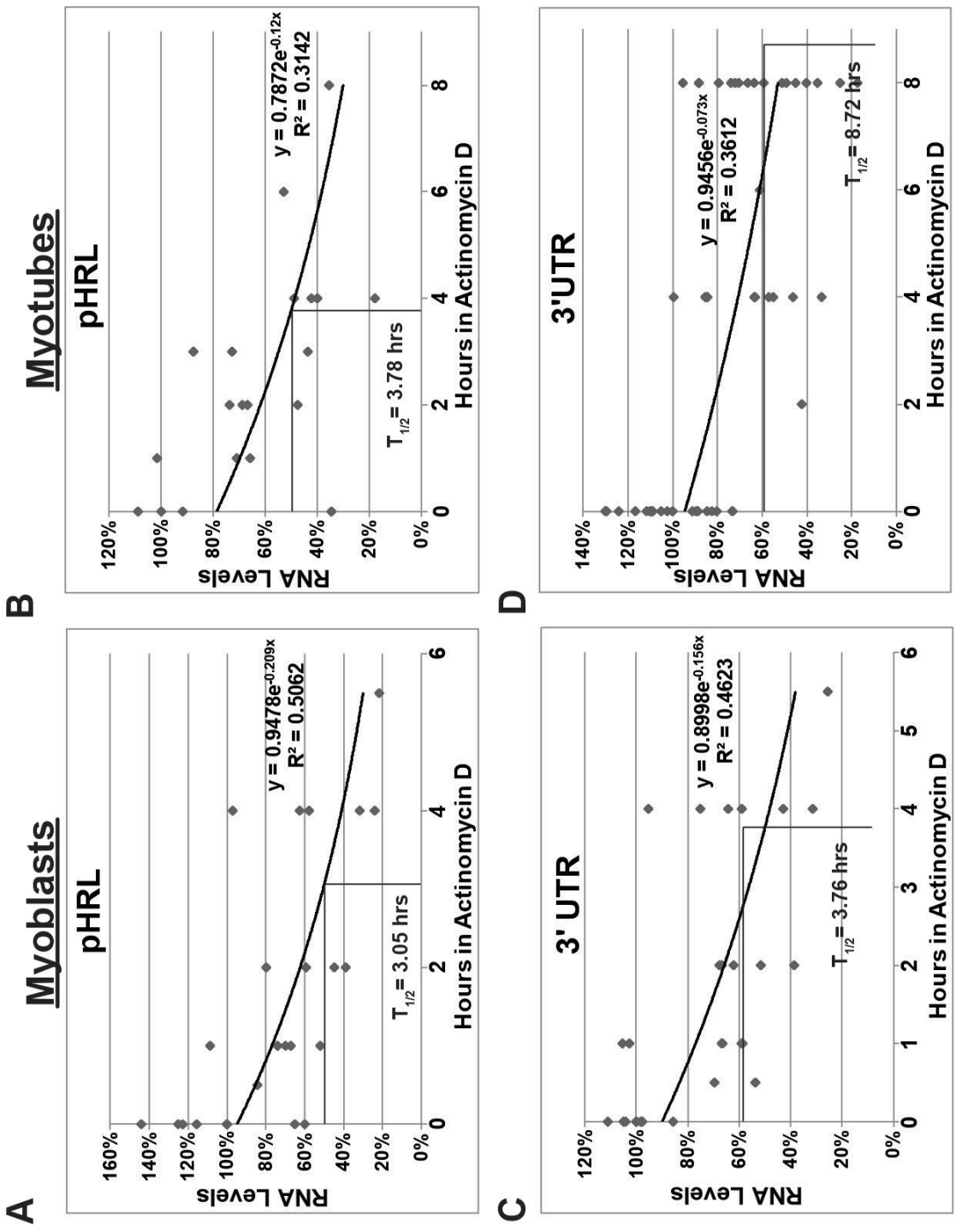
Results

Having shown that the *DMD* 3' UTR increases translation and mRNA abundance in C2C12 myotubes, we were interested in determining what effect the *DMD* 3' UTR has on mRNA stability during differentiation. To do this, we first transfected C2C12 myoblasts with the Renilla pHRL vector (pHRL) or the Renilla *DMD* 3' UTR construct

(3' UTR) we had previously made by replacing the SV40 poly A region of the pHRL vector with the entire *DMD* 3' UTR. After 2 days, we measured Renilla mRNA stability by replacing the growth media of the C2C12 cells with media containing the transcription inhibitor, Actinomycin D, and used qPCR to measure mRNA levels of the transfected constructs compared to total RNA at various time points after Actinomycin D addition (Figure A.1A and A.1C, Myoblasts). To measure mRNA stability of the pHRL and *DMD* 3' UTR constructs in C2C12 myotubes, transfected myoblasts were placed in a low serum media for 5 days to induce differentiation before Actinomycin D was added and mRNA levels were measured using qPCR at various time points (Figure A.1B and A.1D, Myotubes). Because the half life of the Renilla constructs was unknown, we duplicated the experiment in at least four independent experiments for each transfected construct and measured mRNA levels at time points between 0 and 17 hours after Actinomycin D exposure. The results of all independent experiments were combined and the half life was estimated using a nonlinear regression analysis to plot the best fit line (Figure A.1). The estimated half life of the control pHRL construct (pHRL) and the *DMD* 3'UTR construct (3' UTR) was similar in C2C12 myoblasts ($T_{1/2}$ = 3.05 hours and 3.76 hours, respectively) (Figure A.1A and A.1C). In C2C12 myotubes, the estimated half life of the control pHRL vector (pHRL) increased slightly from 3.05 hours to 3.78 hours (Figure A.1B). However, the construct containing the *DMD* 3' UTR was more stable in myotubes with an estimated half life of 8.72 hours (Figure A.1D, Myotubes). We found that we had a high well to well variation when transfecting C2C12 cells due to differences in differentiation, cell survival, and transfection efficiency (see Appendix C), which made it difficult to detect changes in mRNA stability by comparing mRNA levels

Figure A.1. Half life of the *DMD* 3' UTR construct.

A. RNA levels of the pHRL Renilla construct in transfected C2C12 myoblasts are shown. Data points represent individual transfections of C2C12 cells across at least four independent experiments, and nonlinear regression analysis was used to estimate the half life of the transfected construct.
B. RNA levels of the pHRL Renilla construct in transfected C2C12 myotubes are shown. Data points represent individual transfections of C2C12 cells across at least four independent experiments, and nonlinear regression analysis was used to estimate the half life of the transfected construct. The estimated half life ($T_{1/2}$) of pHRL Renilla mRNA was similar in both myoblasts and myotubes (3.05 hours and 3.78 hours, respectively).
C. RNA levels of the *DMD* 3' UTR construct in transfected C2C12 myoblasts are shown. Data points represent individual transfections of C2C12 cells across at least four independent experiments, and nonlinear regression analysis was used to estimate the half life of the transfected construct.
D. RNA levels of the *DMD* 3' UTR construct in transfected C2C12 myotubes are shown. Data points represent individual transfections of C2C12 cells across at least four independent experiments, and nonlinear regression analysis was used to estimate the half life of the transfected construct. The *DMD* 3'UTR construct was more stable in C2C12 myotubes compared to myoblasts ($T_{1/2}$ = 8.72 hours and 3.76 hours, respectively). RNA levels were measured using qPCR and plotted as the percent of RNA remaining normalized to the average amount of RNA measured before Actinomycin D treatment (0 hours) in each experiment.


A
Myoblasts
pHRL
y = 0.9478e-0.209x
R² = 0.5062
T1/2 = 3.05 hrs
RNA Levels
Hours in Actinomycin D
B
Myotubes
pHRL
y = 0.7872e-0.12x
R² = 0.3142
T1/2 = 3.78 hrs
RNA Levels
Hours in Actinomycin D
C
3' UTR
y = 0.8998e-0.156x
R² = 0.4623
T1/2 = 3.76 hrs
RNA Levels
Hours in Actinomycin D
D
3'UTR
y = 0.9456e-0.073x
R² = 0.3612
T1/2 = 8.72 hrs
RNA Levels
Hours in Actinomycin D

to total RNA. To overcome this, we designed a more accurate method for detecting changes in mRNA stability that would take into account any well to well variation in transfecting C2C12 cells. Specifically, we designed a 3' UTR tagged construct (3'UTR-tag) that was identical to our original Renilla *DMD* 3' UTR construct (3'UTR) except with the addition of a random 18 bp sequence inserted between the Renilla coding sequence and the *DMD* 3' UTR. This allowed us to design qPCR primers that could selectively amplify the 3'UTR-tag construct containing the full length *DMD* 3' UTR and any 3' UTR deletion construct cotransfected with this construct. To determine what effect the conserved regions of the *DMD* 3' UTR has on mRNA stability, we transfected C2C12 myoblasts with either the full length *DMD* 3' UTR Renilla construct (3'UTR) or the deletion constructs we had previously made spanning the Lemaire A region (dl-A), the middle region of the 3' UTR (dl-M), or the Lemaire D region (dl-D) (See Chapter 1 for deletion boundaries). All cells were cotransfected with the 3'UTR tagged construct (3'UTR-tag) containing the full length 3'UTR. After 2 days in growth media for myoblasts or an additional 5 days in differentiation media for myotubes, Actinomycin D was added and mRNA levels were measured for the 3' UTR construct and the cotransfected 3'UTR tagged construct using qPCR after 0 hours, 4 hours, or 8 hours after Actinomycin D exposure (Figure A.2, Figure A.3). In myoblasts, the mRNA levels of the full length 3'UTR construct and the deletion constructs with the Lemaire A region (dl-A), the middle region (dl-M), or the Lemaire D region (dl-D) deleted remained the same as the cotransfected 3'UTR tagged construct (3'UTR-tag) after 8 hours in Actinomycin D, showing the mRNA decay rate remains the same when any of these regions are deleted in myoblasts (Figure A.2). However, in myotubes there was a ~ 20%

Figure A.2. Relative half life of *DMD* 3' UTR constructs in C2C12 myoblasts.

A. Relative RNA levels of the dl-A construct transfected in C2C12 myoblasts treated with Actinomycin D are shown. The construct was cotransfected with the control 3'UTR-tag construct and RNA levels were normalized to the amount of the 3'UTR-tag mRNA at each time point. There was no significant difference between the deletion construct and the cotransfected 3'UTR-tag construct.
B. Relative RNA levels of the dl-M construct transfected in C2C12 myoblasts treated with Actinomycin D are shown. The construct was cotransfected with the control 3'UTR-tag construct and RNA levels were normalized to the amount of the 3'UTR-tag mRNA at each time point. There was no significant difference between the deletion construct and the cotransfected 3'UTR-tag construct.
C. Relative RNA levels of the dl-D construct transfected in C2C12 myoblasts treated with Actinomycin D are shown. The construct was cotransfected with the control 3'UTR-tag construct and RNA levels were normalized to the amount of the 3'UTR-tag mRNA at each time point. There was no significant difference between the deletion construct and the cotransfected 3'UTR-tag construct. There was no significant difference between the mRNA levels of the *DMD* 3' UTR construct (3'UTR) and the cotransfected 3'UTR-tag construct after 8 hours of Actinomycin D treatment showing that the decay rate is the same for both constructs. The average of four biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.

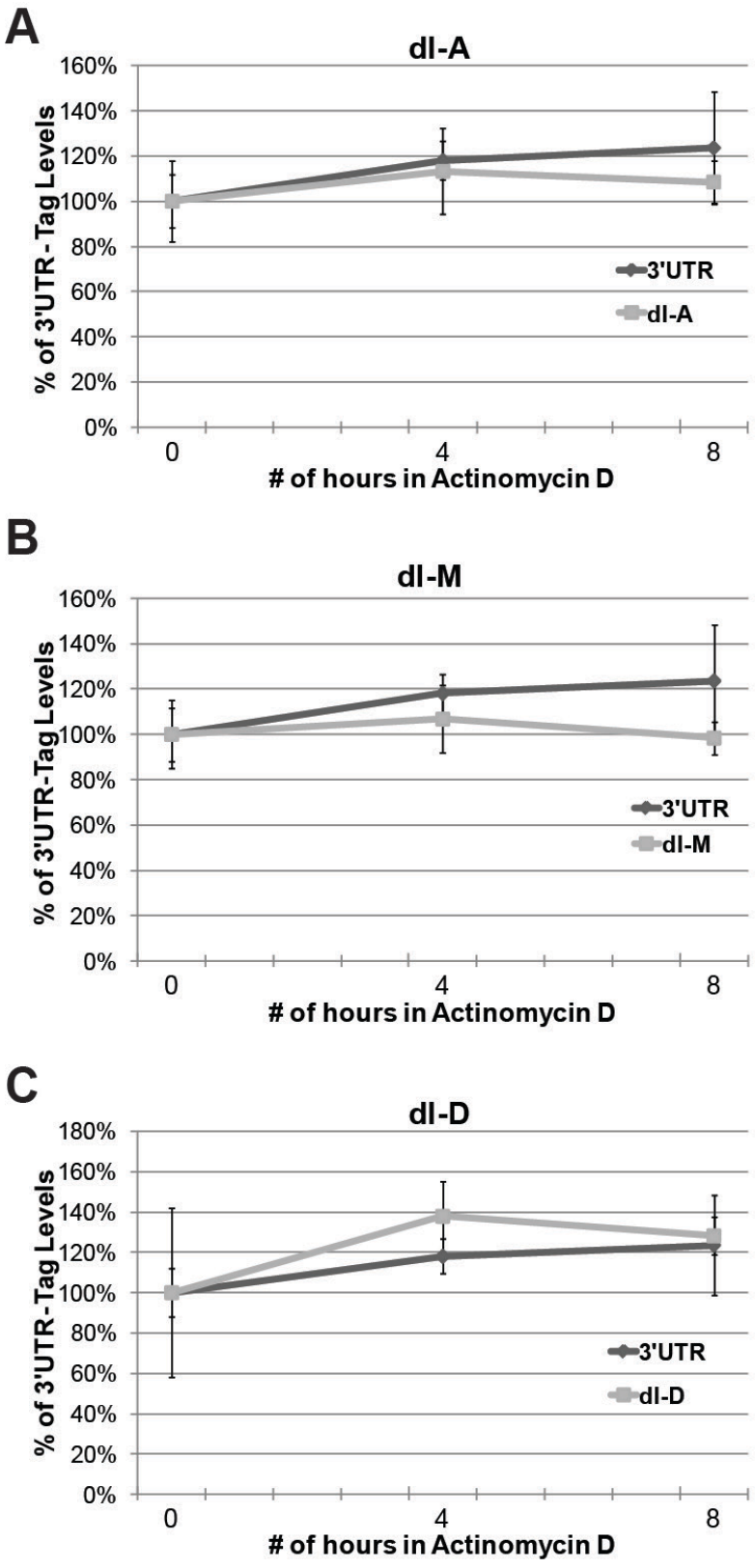
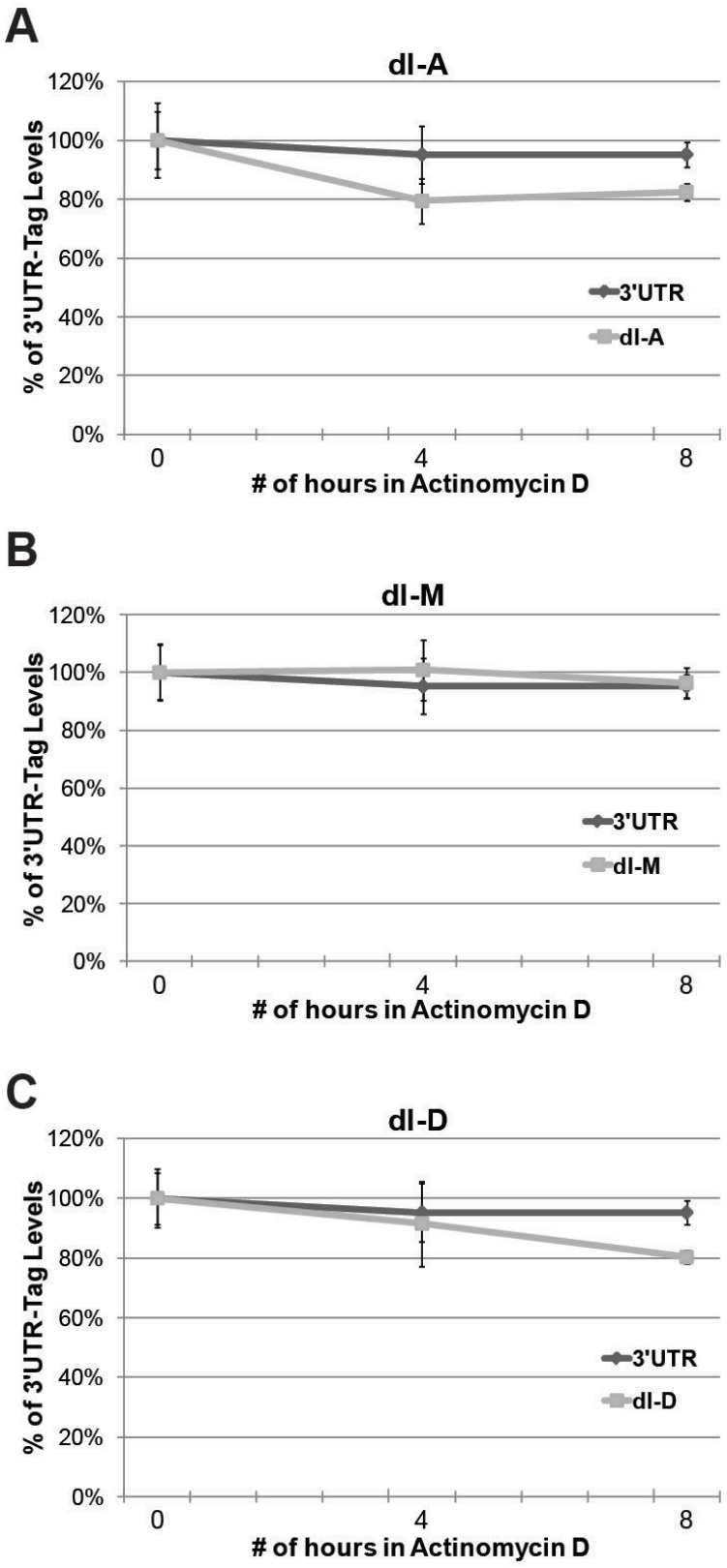

A
dl-A
160%
140%
120%
100%
80%
60%
40%
20%
0%
% of 3'UTR - Tag Levels
3'UTR
dl-A
0
4
8
# of hours in Actinomycin D
B
dl-M
160%
140%
120%
100%
80%
60%
40%
20%
0%
% of 3'UTR-Tag Levels
3'UTR
dl-M
0
4
8
# of hours in Actinomycin D
C
dl-D
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
% of 3'UTR-Tag Levels
3'UTR
dl-D
0
4
8
# of hours in Actinomycin D

Figure A.3. Relative half life of *DMD* 3' UTR constructs in C2C12 myotubes.

A. Relative RNA levels of the dl-A construct transfected in C2C12 myotubes treated with Actinomycin D are shown. The construct was cotransfected with the control 3'UTR-tag construct and RNA levels were normalized to the amount of the 3'UTR-tag mRNA at each time point. There was ~20% less dl-A RNA after 8 hours in Actinomycin D compared to the full length 3' UTR construct.
B. Relative RNA levels of the dl-M construct transfected in C2C12 myotubes treated with Actinomycin D are shown. The construct was cotransfected with the control 3'UTR-tag construct and RNA levels were normalized to the amount of the 3'UTR-tag mRNA at each time point. There was no significant difference between the deletion construct and the cotransfected 3'UTR-tag construct.
C. Relative RNA levels of the dl-D construct transfected in C2C12 myotubes treated with Actinomycin D are shown. The construct was cotransfected with the control 3'UTR-tag construct and RNA levels were normalized to the amount of the 3'UTR-tag mRNA at each time point. There was ~20% less dl-D RNA after 8 hours in Actinomycin D compared to the full length 3' UTR construct. The average of four biological transfection replicates is shown for each construct. Error bars equal +/- 1 standard deviation.


A
dl-A
120%
100%
80%
60%
40%
20%
0%
% of 3'UTR-Tag Levels
3'UTR
dl-A
0
4
8
# of hours in Actinomycin D
B
dl-M
120%
100%
80%
60%
40%
20%
0%
% of 3'UTR-Tag Levels
3'UTR
dl-M
0
4
8
# of hours in Actinomycin D
C
dl-D
120%
100%
80%
60%
40%
20%
0%
% of 3'UTR-Tag Levels
3'UTR
dl-D
0
4
8
# of hours in Actinomycin D

decrease in mRNA levels when the Lemaire A or D regions were deleted (dl-A and dl-D) compared to the cotransfected 3'UTR tagged construct after 8 hours of Actinomycin D exposure showing that the mRNA decay rate was faster when these conserved regions were deleted relative to the full length 3' UTR (Figure A.3).

To measure the actual change in mRNA half life when the conserved Lemaire A region is deleted from the *DMD* 3' UTR in myotubes, we transfected C2C12 myoblasts with the 3' UTR deletion construct with the Lemaire A region removed (dl-A) and differentiated the cells in a low serum media for five days. After differentiation, we added Actinomycin D to the cell media and estimated the mRNA half life of the dl-A construct as previously described (Figure A.4). The estimated half life for the dl-A construct in myotubes is 7.03 hours (Figure A.4) compared to 8.72 hours of the full length 3'UTR construct in C2C12 myotubes (Figure A.1, Myotubes).

Discussion

We had shown previously that the highly conserved elements spanning the Lemaire A and D regions affect mRNA abundance in C2C12 myotubes. Here, we directly measure the effect the *DMD* 3' UTR and these conserved regions have on mRNA stability. Adding the *DMD* 3' UTR to the Renilla pHRL construct increased mRNA half life slightly in C2C12 myoblasts with the 3' UTR construct being 3.76 hours versus 3.05 hours for the pHRL construct, but more than doubled the mRNA half life in C2C12 myotubes with the 3' UTR construct being 8.72 hours versus 3.78 hours for the pHRL construct, showing that the *DMD* 3' UTR stabilizes the mRNA transcript during differentiation. It was difficult to obtain accurate half life measurements of transfected constructs in C2C12 cells, mostly due to well to well variation in C2C12 cell

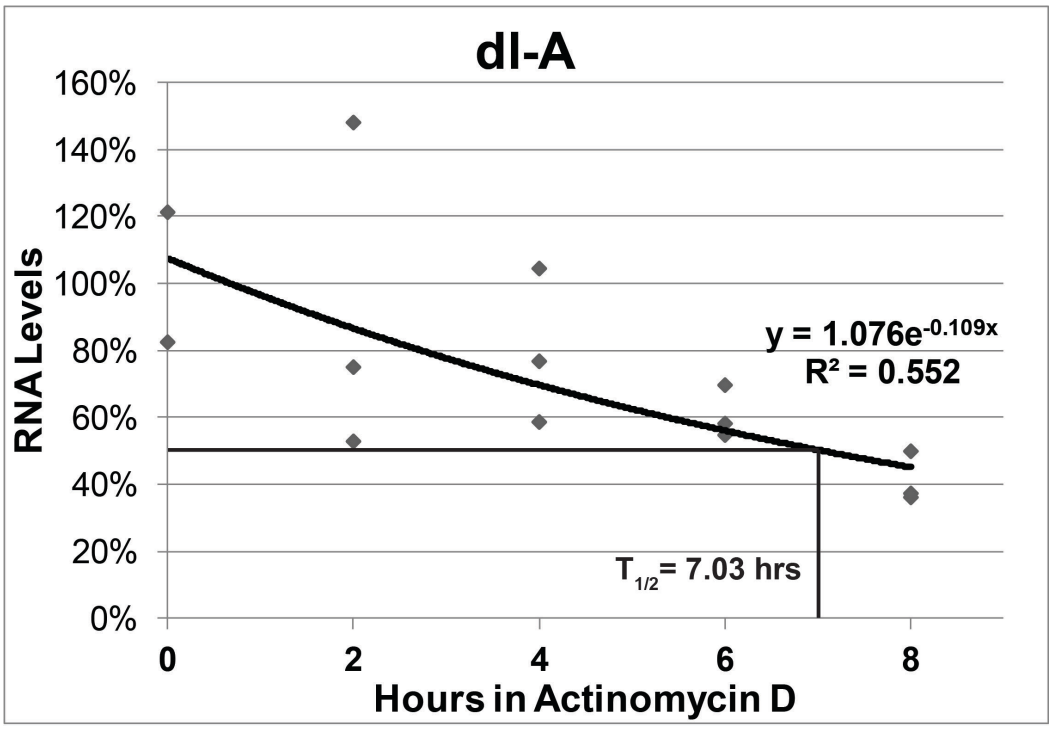


Figure A.4. Half life of the dl-A construct.

RNA levels of the *DMD* 3' UTR Renilla construct with the Lemaire A region deleted (dl-A) in transfected C2C12 myotubes treated with Actinomycin D is shown. Data points represent individual transfections of C2C12 cells across two independent experiments. Nonlinear regression analysis was used to estimate the half life of the dl-A construct in C2C12 myotubes. The estimated half life ($T_{1/2}$) of the dl-A construct is 7.03 hours. RNA levels were measured using qPCR and plotted as the percent of RNA remaining normalized to the average amount of RNA measured before Actinomycin D treatment (0 hours) in each experiment.

differentiation and transfection efficiency (see Appendix C). A cotransfected control that would account for this variation would degrade when the cells were exposed to Actinomycin D and could not be used. To overcome this, we cotransfected our deletion constructs with a modified 3' UTR construct (3'UTR-tag). Although we could not directly measure mRNA half life of each deletion construct, we could more accurately determine whether deleting a region of the 3' UTR had an effect on mRNA stability compared to the full length 3' UTR because all measurements could be normalized to the cotransfected control to account for any well to well variation. When we deleted the Lemaire A region (dl-A), the middle portion of the 3' UTR (dl-M), or the Lemaire D region (dl-D), we did not see a change in mRNA stability in myoblasts (Figure A.2). However, in myotubes, we saw a decrease in mRNA stability when the conserved Lemaire A or D regions were deleted (dl-A and dl-D) (Figure A.3), with 20% less RNA measured for these deletion constructs compared to the full length 3' UTR construct after 8 hours in Actinomycin D. Assuming the half life of the full length 3' UTR construct is 8.72 hours in myotubes, the calculated half life of the dl-A construct would be 7.11 +/- 0.59 hours (Elkon et al., 2010; Perez-Ortin et al., 2007). In a separate experiment, we measured the half life of the dl-A construct in C2C12 myotubes at 7.03 hours (Figure A.4). According to mRNA kinetics calculations (Perez-Ortin et al., 2007), this 19% decrease in half life would result in a 19% decrease in steady state mRNA levels as long as transcription rate remains constant. Indeed, we measured a 20% decrease in steady state RNA levels in C2C12 myotubes when the Lemaire A region was deleted (dl-A) in the cells that were used in these experiments. It should be noted that this decrease of 20% in mRNA levels is smaller than the decrease we showed previously where deleting

the Lemaire A region of the 3' UTR decreased mRNA levels by ~50% in C2C12 myotubes (Chapter 1). This is likely due to differences in C2C12 differentiation states between the two experiments and is discussed more fully in Appendix C. Although we did get a large amount of variation using no transfection control when measuring the half life of these constructs, the separate experiments measuring the estimated half life, the relative half life using the cotransfected 3'UTR-tag construct, and measuring steady state mRNA levels all agree. We conclude that these regions do affect mRNA stability, but the half life of our reporter constructs is dependent on how well the cells are differentiated.

Methods

Cell culture

Mouse C2C12 myoblast cells were purchased from ATCC (CRL-1772). C2C12 myoblasts were cultured in growth medium consisting of Dulbecco's modified Eagle's medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Hyclone). Myogenic differentiation was induced by replacing the medium of the C2C12 myoblasts with DMEM/F-12 media supplemented with 2% horse serum (Sigma) for 5 days. Cells were cultured at 37°C and 5% CO_2.

Transfecting C2C12 cells

When mRNA levels of transfected constructs were analyzed, C2C12 myoblasts ($6x10^5$ cells) were transfected with 275 ng of the construct (137.5 ng of each construct if two constructs were cotransfected) using 3.3 uL of Lipofectamine 2000 in a 24 well tissue culture plate (Falcon) with 2 cm^2 growth area, and grown for 2 days in growth

media.

Measuring mRNA stability

Half lives of transfected constructs were determined by culturing transfected C2C12 cells in Actinomycin D (5µg/mL) at 37°C to inhibit transcription. No cell death or changes in morphology were observed when the cells were exposed to Actinomycin D for up to 8 hours. Total RNA was extracted from C2C12 cells at the times indicated in each figure using Trizol (Invitrogen) as described by the manufacturer. Each RNA sample was treated with DNAse I (Thermo Scientific) to remove any DNA contamination, and cDNA was synthesized using the SuperScript III First Strand Synthesis Kit (Life Technologies) using random hexamers. Real time PCR was performed using the Applied Biosystems 7900HT Real time PCR system using Express SYBR GreenER qPCR SuperMix (Life Technologies) with primers optimized to amplify each construct according to the manufacturer's protocols.

When a single Renilla construct was transfected into C2C12 cells, 500 ng of total RNA was added to each qPCR reaction and the primers GGCTTGTCTGGCCTTTCACT and CCCTTCACCTTCACGAACTC were used to amplify all Renilla constructs used as described above. Results were analyzed using SDS 2.3 software and calculated using the comparative C_T ($\Delta\Delta C_T$) method by normalizing the transfected construct levels at each time point to the average amount of RNA measured for that construct before Actinomycin D treatment (t=0) set at 100%. The percent mRNA remaining was plotted versus time of Actinomycin D addition by combining the data of at least four independent experiments, and nonlinear regression analysis was done using Excel (Microsoft, Redmond, WA) to calculate the best fit line for the data and estimate the half life of the

construct.

To measure the relative stability of a deletion construct, a full length *DMD* 3' UTR construct modified to include an 18 bp insertion between the Renilla coding sequence and the *DMD* 3' UTR (3'UTR-tag) was cotransfected with a deletion construct. This 3'UTR tagged construct was made using the Phusion Site Directed Mutagenesis Kit (Thermo Scientific), as specified by the manufacturer using the *DMD* 3' UTR construct we had previously made as a template and the phosphorylated primers CGGTAAGAAGGAGGAAGTCTTTTCCACATG and GTCCAAGTCCAGAATTACTGCTCGTTCTTCAGC. This resulted in the inclusion of the sequence GGACTTGGACCGGTAAGA in the construct. The primers CTGAGGAGTTCGCTGCCTAC and CTCTTACCGGTCCAAGTCC were used during qPCR to selectively amplify the 3'UTR tagged construct. When the 3'UTR tagged construct was cotransfected with other Renilla constructs, the primers CTGAGGAGTTCGCTGCCTAC and TGTGGAAAAGACTTCCTCCTAGAA were used to selectively amplify the 3'UTR, dl-M, and dl-D constructs. The primers CTGAGGAGTTCGCTGCCTAC and CTGACAGTTCTCAAATGCCTAGAA were used to selectively amplify the dl-A construct. Results were analyzed using SDS 2.3 software and calculated using the comparative C_T ($\Delta\Delta C_T$) method by normalizing each mRNA level to the cotransfected 3'UTR-tag construct mRNA at each time point.

References

Andreassi, C., and Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. Trends Cell Biol *19*, 465-474.

Elkon, R., Zlotorynski, E., Zeller, K.I., and Agami, R. (2010). Major role for mRNA stability in shaping the kinetics of gene induction. BMC Genomics *11*, 259.

Gramolini, A.O., Belanger, G., and Jasmin, B.J. (2001). Distinct regions in the 3' untranslated region are responsible for targeting and stabilizing utrophin transcripts in skeletal muscle cells. J Cell Biol *154*, 1173-1183.

Matoulkova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol *9*, 563-576.

Mazumder, B., Seshadri, V., and Fox, P.L. (2003). Translational control by the 3'-UTR: the ends specify the means. Trends Biochem Sci *28*, 91-98.

Perez-Ortin, J.E., Alepuz, P.M., and Moreno, J. (2007). Genomics and gene transcription kinetics in yeast. Trends Genet *23*, 250-257.

Spitali, P., van den Bergen, J.C., Verhaart, I.E., Wokke, B., Janson, A.A., van den Eijnde, R., den Dunnen, J.T., Laros, J.F., Verschuuren, J.J., t Hoen, P.A.*, et al.* (2013). DMD transcript imbalance determines dystrophin levels. FASEB J *27*, 4909-4916.

APPENDIX B

CHARACTERIZATION OF HUMAN *DMD* 3' UTR VARIANTS AND PREDICTED BINDING SITES

3' UTR variants in human DMD patients

The sequence of human *DMD* 3' UTRs from a population of 1,222 patients was obtained from Dr. Robert Weiss (Department of Human Genetics, University of Utah). This group of patients included 410 patients enrolled in the United Dystrophinopathy Project (UDP), 569 patients from clinical referrals of doctors who had requested *DMD* mutation screening for individuals suspected to be dystrophic, and 243 patients from a newborn screening study that screened for *DMD* mutations in infants with high creatine kinase (CK) levels. There were a total of 1,843 discrepancies compared to a reference sequence identified amongst 626 3' UTR sequences using Consed. Each discrepancy was visually inspected to determine whether the change was due to a sequencing error or represented a true variant. There were 1,377 discrepancies amongst 435 3' UTRs that were verified as true variants. Identified variants were named using the HGVS guidelines and recommendations for variant nomenclature (den Dunnen and Antonarakis, 2000) (Table B.1). There were 787 3' UTRs that matched the reference sequence and contained no variants. Twenty-three variants in the *DMD* 3' UTR were found only in patients labeled as having no mutations in the coding region of the *DMD* gene (NODCM) (Table B.1). Based on our deletion constructs made earlier (Chapter 1), the 3' UTR region of

Table B.1

Variants found in the *DMD* 3' UTR of human patients.

#	Variant	# of Hits	3' UTR Position		Mut. Type	Variant Code	3'UTR Region	Phast Cons Score	All NoDCM
1	Sequencing Error	355	Other	0					
2	c.11047-249A>T	136	Intron	0	SNP	A			
3	No Trace Information available	110	Other	0					
4	c.11047-150_11047-149insA	9	Intron	0	INDEL	B			
5	c.11047-389_11047-386dupATCA	2	Intron	0	INDEL	M1			
6	c.11047-255T>G	1	Intron	0	SNP	M2			
7	c.11047-300_11047-299insT	1	Intron	0	INDEL	M3			NODCM
8	c.11047-377_11047-376insC	1	Intron	0	INDEL	M4			
9	c.11047-389_11047-388insATCAA	1	Intron	0	INDEL	M5			
10	c.11047-451_11047-450insG	1	Intron	0	INDEL	M6			NODCM
11	c.11047-85G>T	1	Intron	0	SNP	M7			
12	c.*38G>A	4	3UTR	38	SNP	M8	Aa	0.99	
13	c.*64_*65insC	2	3UTR	64	INDEL	M9	Aa	1.00	NODCM
14	c.*93_*94insT	4	3UTR	93	INDEL	M10	Aa	1.00	
15	c.*266C>T	1	3UTR	266	SNP	M11	Ab	1.00	
16	c.*428G>C	1	3UTR	428	SNP	M12	Ac	1.00	
17	c.*467_*476del	4	3UTR	467	INDEL	M13	Ad	0.35	
18	c.*471_*476delCACACA	1	3UTR	471	INDEL	M14	Ad	0.89	
19	c.*475_*476dupCA	3	3UTR	475	INDEL	M15	Ad	0.42	
20	c.*477_*480dupTACA	214	3UTR	477	INDEL	C	Ad	0.00	
21	c.*477_*484del	24	3UTR	477	INDEL	D	Ad	0.00	
22	c.*477_*490del	9	3UTR	477	INDEL	E	Ad	0.00	
23	c.*477_*488del	5	3UTR	477	INDEL	F	Ad	0.00	
24	c.*477_*480delTACA	2	3UTR	477	INDEL	M16	Ad	0.00	
25	c.*477_*482delTACACA	2	3UTR	477	INDEL	M17	Ad	0.00	
26	c.*477_*486del	1	3UTR	477	INDEL	M18	Ad	0.00	
27	c.*477T>C	1	3UTR	477	SNP	M19	Ad	0.00	NODCM
28	c.*480_*481insTA	2	3UTR	480	INDEL	M20	Ad	0.05	
29	c.*480_*481insTACACA	1	3UTR	480	INDEL	M21	Ad	0.05	NODCM
30	c.*483_*492del	1	3UTR	483	INDEL	M22	Ad	0.11	
31	c.*489C>T	4	3UTR	489	SNP	M23	Ad	0.99	
32	c.*489_*492dupCACA	1	3UTR	489	INDEL	M24	Ad	0.99	
33	c.*491_*492delCA	5	3UTR	491	INDEL	G	Ad	0.87	
34	c.*491_*492dupCA	2	3UTR	491	INDEL	M25	Ad	0.87	
35	c.*571A>G	1	3UTR	571	SNP	M26	Ad	1.00	NODCM
36	c.*651G>A	1	3UTR	651	SNP	M27	Ma	0.94	NODCM
37	c.*755_*758dupATTT	1	3UTR	755	INDEL	M28	Ma	0.00	NODCM
38	c.*809C>T	4	3UTR	809	SNP	M29	Ma	0.28	
39	c.*921_*924dupACAA	1	3UTR	921	INDEL	M30	Ma	0.65	
40	c.*952_*956delAAGTA	2	3UTR	952	INDEL	M31	Ma	0.00	
41	c.*952_*953dupAA	1	3UTR	952	INDEL	M32	Ma	0.00	
42	c.*968_*969insT	3	3UTR	968	INDEL	M33	Ma	0.00	
43	c.*998_*1001dupCTTA	2	3UTR	998	INDEL	M34	Ma	0.96	
44	c.*1006_*1009dupATCA	1	3UTR	1006	INDEL	M35	Ma	0.00	
45	c.*1022_*1023insATTG	1	3UTR	1022	INDEL	M36	M	0.00	
46	c.*1031_*1038del	4	3UTR	1031	INDEL	M37	Mb	0.06	
47	c.*1031_*1042del	1	3UTR	1031	INDEL	M38	Mb	0.06	NODCM
48	c.*1031G>A	1	3UTR	1031	SNP	M39	Mb	0.06	
49	c.*1045_*1052dupTGATTGAT	3	3UTR	1045	INDEL	M40	Mb	0.01	

Table B.1 continued

#	Variant	# of Hits	3' UTR Position		Mut. Type	Variant Code	3'UTR Region	Phast Cons Score	All NoDCM
50	c.*1046_*1048dupGAT	6	3UTR	1046	INDEL	H	Mb	0.01	
51	c.*1049_*1052dupTGAT	201	3UTR	1049	INDEL	I	Mb	0.06	
52	c.*1049_*1052delTGAT	18	3UTR	1049	INDEL	J	Mb	0.06	
53	c.*1052_*1053insTGGTTGATTGATTGATTGATTGAT	1	3UTR	1052	INDEL	M41	Mb	0.04	NODCM
54	c.*1052_*1053insTGGTTGATTGATTGATTGATTGAT	1	3UTR	1052	INDEL	M42	Mb	0.04	NODCM
55	c.*1165C>T	1	3UTR	1165	SNP	M43	Mb	0.00	
56	c.*1185_*1188dupTGAT	5	3UTR	1185	INDEL	K	Mb	0.00	
57	c.*1242_*1245dupCTTT	1	3UTR	1242	INDEL	M44	Mb	0.02	
58	c.*1285C>A	6	3UTR	1285	SNP	L	Mb	0.00	
59	c.*1353_*1354insT	1	3UTR	1353	INDEL	M45	Mb	0.01	
60	c.*1393T>A	1	3UTR	1393	SNP	M46	Mb	0.02	
61	c.*1401C>A	3	3UTR	1401	SNP	M47	Mb	0.52	
62	c.*1447A>G	312	3UTR	1447	SNP	N	Mb	0.19	
63	c.*1476G>A	1	3UTR	1476	SNP	M48	Mb	0.28	NODCM
64	c.*1533G>A	1	3UTR	1533	SNP	M49	Mc	0.02	
65	c.*1562_*1565dupTAAG	67	3UTR	1562	INDEL	O	Mc	0.01	
66	c.*1588_*1591dupTGTT	7	3UTR	1588	INDEL	P	Mc	0.00	
67	c.*1600_*1601insA	1	3UTR	1600	INDEL	M50	Mc	0.00	NODCM
68	c.*1638A>G	5	3UTR	1638	SNP	Q	Mc	0.00	
69	c.*1705A>G	1	3UTR	1705	SNP	M51	Mc	1.00	
70	c.*1716_*1717insC	1	3UTR	1716	INDEL	M52	Mc	0.99	
71	c.*1731A>C	1	3UTR	1731	SNP	M53	Mc	0.01	
72	c.*1750_*1751insA	1	3UTR	1750	INDEL	M54	Mc	0.00	
73	c.*1758_*1761delTTCA	4	3UTR	1758	INDEL	M55	Mc	0.00	
74	c.*1783A>G	9	3UTR	1783	SNP	R	Mc	0.86	
75	c.*1799G>T	9	3UTR	1799	SNP	S	Mc	1.00	
76	c.*1838_*1841delTGAT	1	3UTR	1838	INDEL	M56	Mc	0.99	NODCM
77	c.*1903_*1906dupTTAA	2	3UTR	1903	INDEL	M57	Mc/Md	0.99	
78	c.*1951_*1952insT	2	3UTR	1951	INDEL	M58	Md	1.00	
79	c.*2018_*2021dupCTTT	1	3UTR	2018	INDEL	M59	Md	0.01	NODCM
80	c.*2019T>C	1	3UTR	2019	SNP	M60	Md	0.00	NODCM
81	c.*2028_*2048del	1	3UTR	2028	INDEL	M61	Md	0.30	
82	c.*2030_*2033dupTAAT	1	3UTR	2030	INDEL	M62	Md	0.00	
83	c.*2031_*2034dupAATT	1	3UTR	2031	INDEL	M63	Md	0.01	
84	c.*2037_*2038dupTT	1	3UTR	2037	INDEL	M64	Md	0.29	NODCM
85	c.*2039C>A	4	3UTR	2039	SNP	M65	Md	0.34	
86	c.*2103T>A	1	3UTR	2103	SNP	M66	Md	0.01	
87	c.*2155_*2156insATGTGACGCTGGACCTT	4	3UTR	2155	INDEL	M67	Md	0.75	
88	c.*2234_*2237dupAAGT	32	3UTR	2234	INDEL	T	Da	0.98	
89	c.*2245T>C	1	3UTR	2245	SNP	M68	Da	1.00	NODCM
90	c.*2268C>T	1	3UTR	2268	SNP	M69	Da	1.00	NODCM
91	c.*2314T>C	1	3UTR	2314	SNP	M70	Da	1.00	
92	c.*2326G>A	1	3UTR	2326	SNP	M71	Da	1.00	NODCM
93	c.*2548_*2549insC	1	3UTR	2548	INDEL	M72	Db	1.00	
94	c.*2578_*2581dupACTT	1	3UTR	2578	INDEL	M73	Db	1.00	

Table B.1 continued

#	Variant	# of Hits	3' UTR Position		Mut. Type	Variant Code	3'UTR Region	Phast Cons Score	All NoDCM
95	c.*2591T>A	1	3UTR	2591	SNP	M74	Dc	1.00	NODCM
96	c.*2701T>G	2	3UTR	2701	SNP	M75	PolyA region	0.00	
97	c.*2722_*2727delGTTTTT	1	3UTR	2722	INDEL	M76	PolyA region	0.05	NODCM
98	c.*2743T>C	2	3UTR	2743	SNP	M77	PolyA region	0.02	
99	c.*2760_*2763delACTT	156	3UTR	2760	INDEL	U	PolyA region	0.01	
100	c.*2760_*2763dupACTT	2	3UTR	2760	INDEL	M78	PolyA region	0.01	NODCM
101	c.*2768_*2771dupAATA	13	3UTR	2768	INDEL	V	PolyA region	0.00	
102	c.*2774_*2777dupTAAT	1	3UTR	2774	INDEL	M79	PolyA region	0.00	
103	c.*2819A>C	1	3UTR	2819	SNP	M80	PolyA region	0.00	NODCM

Note: Variants found in human *DMD* 3' UTRs are shown. '# of Hits' equals the number of 3' UTRs each variant was found in. The 3' UTR position for each variant was determined where nucleotide #1 is the nucleotide immediately after the stop codon. If the variant affected more than one nucleotide, the position of the first nucleotide is given. A 'Variant Code' was assigned for each variant as a shorthand way of referring to each variant. Any variant found in less than 5 3' UTRs was assigned a code starting with 'M' and followed by a number. Any variant found in 5 or more 3' UTRs was assigned all other letters. Conservation of each variant was determined by the PhastCons score of that position of the 3' UTR extracted from the UCSC genome browser. The PhastCons score is a measurement of conservation indicated by a score between 0 (not conserved) and 1 (conserved). PhastCons scores above 0.85 are highlighted. One hundred and one unique variants were identified. "Sequencing Error" and "No Trace Information Available" represent the number of times a discrepancy between the sequence and a reference sequence was miscalled by Consed or where the sequence trace was inaccessible, respectively.

each variant was determined (Table B.1). The conservation of each variant was determined by the measured PhastCons score extracted from the UCSC Genome Browser (Table B.1). One hundred and one total variants were identified in the *DMD* 3' UTR and surrounding regions of our patient population and given a variant code; where rare variants (found in less than five patients) were given a label M followed by a #, and more common variants (found in five or more patients) were coded with all other letters (Table B.1). The haplotype of all patients containing at least one variant is shown (Table B.2). In addition, we searched for *DMD* 3' UTR variants in publicly available databases, and found 83 variants in the current version of dbSNP (Sherry et al., 2001), 40 variants in the 1000 Genome database (Abecasis et al., 2012), and 38 variants reported in the Leiden Database (Aartsma-Rus et al., 2006) (Table B.3). One hundred and forty total variants in the *DMD* 3' UTR were identified in our patient population and all databases searched (Table B.3). Predicted miRNA binding sites in the *DMD* 3' UTR were identified using TargetScan (Lewis et al., 2005) (Table B.4). Predicted RNA protein binding sites in the Lemaire A region (Table B.5), a conserved AU rich sequence from the middle region (Table B.6), and the Lemaire D region (Table B.7) of the *DMD* 3' UTR were identified using the RNA Binding Protein DataBase (RBPDB) (Cook et al., 2011).

Table B.2

Haplotype of human patients with 3' UTR variants.

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
1	DOBS059	A	C	I	N	U						NODCM
2	42458	A	C	I	N	U						MISSENSE
3	42514	A	C	I	N	U						MISSENSE
4	DC0110	A	C	I	N	U						STOP
5	DC0246	A	C	I	N	U						NODCM
6	DC0308	A	C	I	N	U						DELETION
7	DC0342	A	C	I	N	U						STOP
8	DC0396	A	C	I	N	U						SPLICE
9	DC0520	A	C	I	N	U						DELETION
10	DC0634A	A	C	I	N	U						SPLICE
11	DC0680	A	C	I	N	U						NODCM
12	DC0867	A	C	I	N	U						STOP
13	DC1094	A	C	I	N	U						NODCM
14	DC1109	A	C	I	N	U						SUBEXONIC
15	DC1129	A	C	I	N	U						STOP
16	DOBS0145	A	C	I	N	U						NODCM
17	DOBS0146	A	C	I	N	U						NODCM
18	DOBS0161	A	C	I	N	U						NODCM
19	DOBS036	A	C	I	N	U						NODCM
20	DOBS083	A	C	I	N	U						NODCM
21	DR0188	A	C	I	N	U						OTHER
22	DR0193	A	C	I	N	U						SPLICE
23	DR43184	A	C	I	N	U						STOP
24	DR43219	A	C	I	N	U						STOP
25	DR43311	A	C	I	N	U						SUBEXONIC
26	DR43497	A	C	I	N	U						MISSENSE
27	DR43567	A	C	I	N	U						DELETION
28	DC1167	A	C	I	N	U						NODCM
29	DC1193	A	C	I	N	U						STOP
30	43020	A	C	I	N	U						DUPLICATION
31	DC0550	A	C	I	N	U						NODCM
32	OSU317	A	C	I	N	U						NODCM
33	DC0551	A	C	M35	I	N	U					DELETION
34	DC0244	A	C	I	N	M55	U					DELETION
35	DOBS014	A	C	I	N	M55	U					NODCM
36	DOBS0201	A	C	M15	I	N	U					NODCM
37	DC0248	A	C	I	N	O	U					NODCM
38	DC0271	A	C	I	N	O	U					NODCM
39	DC0336	A	C	I	N	O	U					NODCM
40	DC0340	A	C	I	N	O	U					NODCM
41	DC0689	A	C	I	N	O	U					NODCM
42	DC0947	A	C	I	N	O	U					NODCM
43	DR43190	A	C	I	N	O	U					DELETION
44	DR43371	A	C	I	N	O	U					DUPLICATION
45	DR43460	A	C	I	N	O	U					DUPLICATION
46	DR43694	A	C	I	N	O	U					SUBEXONIC
47	DR43871	A	C	I	N	O	U					DUPLICATION
48	OSU346	A	C	I	N	O	U					STOP

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
49	OSU383	A	C	I	N	O	U					NODCM
50	OSU396	A	C	I	N	O	U					STOP
51	DOBS0189	A	C	I	N	O	U					NODCM
52	DOBS0239	A	C	I	N	O	U					NODCM
53	DOBS0195	A	C	I	N	O	U					NODCM
54	DR43693	A	C	I	N	T	U					DELETION
55	42971	A	C	I	N	T	U					SPLICE
56	42990	A	C	I	N	T	U					DELETION
57	42991	A	C	I	N	T	U					DELETION
58	DC0681	A	C	I	N	T	U					SUBEXONIC
59	DOBS0120	A	C	I	N	T	U					NOTDMD
60	DOBS0157	A	C	I	N	T	U					NODCM
61	DOBS019	A	C	I	N	T	U					NODCM
62	DOBS032	A	C	I	N	T	U					NODCM
63	DR43545	A	C	I	N	T	U					SUBEXONIC
64	DR43650	A	C	I	N	T	U					STOP
65	DR43652	A	C	I	N	T	U					DUPLICATION
66	DR43771	A	C	I	N	T	U					STOP
67	DR43814	A	C	I	N	T	U					STOP
68	DR43915	A	C	I	N	T	U					SUBEXONIC
69	OSU359	A	C	I	N	T	U					NODCM
70	DOBS0180	A	C	I	N	T	U					NODCM
71	DC0590	A	C	I	N	T	U					SPLICE
72	DR43221	A	C	I	N	T	U					PSEUDOEXON
73	43065C	A	C	I	N							NODCM
74	DOBS043	A	C	I	N							NODCM
75	DR43631	A	C	I	N							NODCM
76	DC0142	A	C	I	N							NODCM
77	DC0427	A	C	I	N							NODCM
78	DC0554	A	C	I	N							DUPLICATION
79	DC0901	A	C	I	N							NODCM
80	DC1072	A	C	I	N							NODCM
81	DOBS0115	A	C	I	N							NOTDMD
82	DOBS0141	A	C	I	N							NODCM
83	DR43296	A	C	I	N							NODCM
84	DR43335	A	C	I	N							DUPLICATION
85	DR43581	A	C	I	N							SUBEXONIC
86	DR43767	A	C	I	N							DELETION
87	OSU341	A	C	I	N							STOP
88	DOBS0199	A	C	I	N							NODCM
89	DOBS0194	A	C	I	N							DELETION
90	42968	A	C	I	N	M55						STOP
91	DC0539	A	C	N	U							SPLICE
92	43026	A	C	N	U							SUBEXONIC
93	DC0677	A	C	N	U							SUBEXONIC
94	DC1086	A	C	N	U							SUBEXONIC
95	DOBS0131	A	C	N	U							NODCM
96	DOBS0139	A	C	N	U							NODCM
97	DOBS070	A	C	N	U							NODCM
98	DR43401	A	C	N	U							DUPLICATION
99	DC1233	A	C	N	U							NODCM
100	DC1152	A	C	N	U							SPLICE

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
101	DOBS0219	A	C	N	U							NODCM
102	DC0079	A	C	N	U							SUBEXONIC
103	DC0234	A	C	N	U	M79						STOP
104	DC0859	A	C	N	M74	U						NODCM
105	DOBS077	A	C	M41	N	U						NODCM
106	DC1100	A	C	N	O							SUBEXONIC
107	DC1116	A	C	N	O							NODCM
108	DOBS0144	A	C	N	O							NODCM
109	DR43116	A	C	I	N	O						STOP
110	43080	A	C	I	N	O						NOTDMD
111	DC0346	A	C	I	N	O						SPLICE
112	DC0395	A	C	I	N	O						STOP
113	DC0624	A	C	I	N	O						SPLICE
114	DR43378	A	I	N								SUBEXONIC
115	DC0111	A	I	N	T	U						DUPLICATION
116	DC0129	A	I	N	O	U						DUPLICATION
117	DC0437	A	I	N	U							STOP
118	DR43387	A	I	N	U							NODCM
119	DC0555	A	M20	I	N	U						SPLICE
120	DC1123	A	M21	I	N	U						NODCM
121	DC1036	A	C	N	O	U						NODCM
122	DC0260	A	C	M40	N	O	U					STOP
123	DC0852	A	C	N								STOP
124	DR0010	A	C	N								SPLICE
125	DR0049	A	C	N								NODCM
126	DC1106	A	N	U								SUBEXONIC
127	DR43405	A	N	U								NODCM
128	DR43271	A	C	I	N	T						STOP
129	DR43268	A	C	J	N	U						DUPLICATION
130	DR0184	A	C	U								STOP
131	DOBS041	A	C	I								NODCM
132	DC0737	A	C	I	M50							NODCM
133	DC0380	A	C									NODCM
134	DC0740	A	C									STOP
135	DC1054	A	M20	N								NODCM
136	DOBS0154	A	C	M27	M38	L	N	U				NODCM
137	42960	C	I	N	U							DUPLICATION
138	42966	C	I	N	U							DUPLICATION
139	43084	C	I	N	U							SPLICE
140	DC0472	C	I	N	U							DUPLICATION
141	DC0571	C	I	N	U							NODCM
142	DOBS035	C	I	N	U							NODCM
143	DR43042A	C	I	N	U							SUBEXONIC
144	DR43218	C	I	N	U							SPLICE
145	DR43563	C	I	N	U							SPLICE
146	DRVH1033	C	I	N	U							SUBEXONIC
147	OSU421	C	I	N	U							NODCM
148	DC0035	C	I	N	U							NODCM
149	DC0399	C	I	N	M68	U						NODCM
150	DC0525	C	I	N	O	U						STOP
151	DRVH1111	C	I	N	O	U						NODCM
152	DC0044	C	I	N	O	U						STOP

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
153	DC0606	C	I	N	O	U						SPLICE
154	DC0635	C	I	N	O	U						DUPLICATION
155	DR43165	C	I	N	O	U						NODCM
156	DR43225	C	I	N	O	U						STOP
157	DRVH43503	C	I	N	O	U						NODCM
158	DRVH1035	C	I	N	O	U						DELETION
159	OSU340	C	I	N	O	U						NODCM
160	DC0644	M9	C	I	N	O	U					NODCM
161	DC0033	C	I	N	T	U						SUBEXONIC
162	DOBS082	C	I	N	T	U						NODCM
163	DR43043A	C	I	N	T	U						STOP
164	DR43800	C	I	N	T	U						STOP
165	DR43865	C	I	N	T	U						DUPLICATION
166	OSU357	C	I	N	T	U						SUBEXONIC
167	DR43222	C	I	N	T	U						NODCM
168	DR43812	C	I	N	O							DUPLICATION
169	DC0004	C	I	N	O							NODCM
170	DC0031	C	I	N	O							NODCM
171	DRVH1124	C	I	N	O							STOP
172	DC0783	C	I	N								SPLICE
173	DRVH1122	C	I	N								DUPLICATION
174	43027	C	I	N								DUPLICATION
175	43076	C	I	N								NOTDMD
176	DC0030	C	I	N								STOP
177	DC0292	C	I	N								STOP
178	DC0466	C	I	N								MISSENSE
179	DC0611	C	I	N								NODCM
180	DOBS0103	C	I	N								NOTDMD
181	DOBS053	C	I	N								NODCM
182	DOBS065	C	I	N								NODCM
183	DOBS084	C	I	N								NODCM
184	DR0122	C	I	N								STOP
185	DR43136	C	I	N								SUBEXONIC
186	DR43155	C	I	N								DELETION
187	DR43174	C	I	N								STOP
188	DR43177	C	I	N								SUBEXONIC
189	DR43230	C	I	N								NODCM
190	DR43233	C	I	N								NODCM
191	DR43590	C	I	N								SUBEXONIC
192	DR43820	C	I	N								SPLICE
193	DRVH1141	C	I	N								SPLICE
194	OSU354	C	I	N								NODCM
195	OSU407	C	I	N								STOP
196	OSU448	C	I	N								NODCM
197	DOBS0187	C	I	N								NODCM
198	DC1150	C	I	N								NODCM
199	DC0089	C	I	N								SUBEXONIC
200	DOBS0205	C	I	N	M51							NODCM
201	DC0451	C	I	O								NODCM
202	DR43283	C	I	N	T							DELETION
203	DR43759	C	I	N	T							STOP
204	DR43691	M10	C	I	N	T						DELETION

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
205	DRVH1091	C	N	O								SUBEXONIC
206	DR43143	C	I									DELETION
207	DOBS0112	C	I	M55								NOTDMD
208	DC0564	M11	C									SUBEXONIC
209	DC0426	C	N	U								SUBEXONIC
210	DOBS071	C	N	U								NODCM
211	OSU414	C	N	U								STOP
212	DC0083	C	M36	M39	N	U						STOP
213	DC0118	C	N									STOP
214	DR43810	C	L	N	U							NODCM
215	DOBS054	C	M37	L	N							NODCM
216	DOBS092	C	M37	L	N	U						NODCM
217	DR43719	C	M37	L	N	U						DUPLICATION
218	DC0056	C	M37	L	N	U						NODCM
219	43030	C	J	U								STOP
220	DC0584	C	J	N	O	Q	U					STOP
221	DR43570	C	U									DUPLICATION
222	OSU387	C	I	T								STOP
223	DC0348	I	N	U								STOP
224	DC0101	I	N	U								NODCM
225	DC0177	I	N	U								NODCM
226	DR43804	I	N	U								DELETION
227	DR43295	I	N	U								DUPLICATION
228	DC0445	I	N	O	U							NODCM
229	DR43417	I	N	O	U							DUPLICATION
230	DC0604	I	N	O								DELETION
231	DR43432	I	N	O								DELETION
232	DC0504	M34	I	N	O							NODCM
233	DOBS0202	M34	I	N	O							NODCM
234	DC0351	I	N									NODCM
235	DC0138	I	N									SUBEXONIC
236	DC0358	I	N									STOP
237	DR43153	I	N									STOP
238	DR43486	I	N									STOP
239	DOBS087	M19	M25	M31	I	N	M76					NODCM
240	DR43755	M8	I	N	M58	M75						SUBEXONIC
241	DR43756	M8	M22	I	N	M58	M75					STOP
242	DC0220	I										STOP
243	DR43328	I										DUPLICATION
244	DR43392	M16	I									STOP
245	DR43439	I	U									SUBEXONIC
246	DR0254	I	M53	U								STOP
247	DC0206	I	N	R								STOP
248	DC0524	N										SPLICE
249	DC0765	N										STOP
250	DOBS011	N										NODCM
251	DR43135	N										DUPLICATION
252	DR43393	N										DUPLICATION
253	42515	N										MISSENSE
254	42969	N										STOP
255	43016	N										STOP
256	43075	N										NODCM

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
257	43158C	N										MISSENSE
258	DC0067	N										STOP
259	DC0071	N										NODCM
260	DC0094	N										NODCM
261	DC0152	N										NODCM
262	DC0160	N										PSEUDOEXON
263	DC0181	N										SPLICE
264	DC0205	N										NODCM
265	DC0299	N										SPLICE
266	DC0310	N										SUBEXONIC
267	DC0366	N										STOP
268	DC0494	N										DELETION
269	DC0592	N										DELETION
270	DC0636	N										SUBEXONIC
271	DC0937	N										STOP
272	DC1037	N										MISSENSE
273	DC1059	N										NODCM
274	DC1081	N										NODCM
275	DOBS0102	N										NOTDMD
276	DOBS0134	N										NODCM
277	DOBS0142	N										NODCM
278	DOBS0168	N										NODCM
279	DOBS027	N										NODCM
280	DOBS028	N										NODCM
281	DOBS033	N										NODCM
282	DOBS039	N										NODCM
283	DOBS049	N										NODCM
284	DOBS063	N										NODCM
285	DOBS073	N										NODCM
286	DR0054	N										STOP
287	DR43198	N										STOP
288	DR43240	N										DUPLICATION
289	DR43301	N										DUPLICATION
290	DR43314	N										DUPLICATION
291	DR43325	N										NODCM
292	DR43617	N										SUBEXONIC
293	DR43722	N										SUBEXONIC
294	DR43740	N										NODCM
295	DR43816	N										DELETION
296	DR43846	N										SPLICE
297	DR43848	N										STOP
298	DRVH43382	N										DUPLICATION
299	OSU328	N										SUBEXONIC
300	OSU411	N										STOP
301	OSU417	N										SUBEXONIC
302	OSU425	N										NODCM
303	DOBS0197	N										NODCM
304	DOBS0184	N										NODCM
305	DC0403	M33	N	R								NODCM
306	DC0678	M5	N	R								DELETION
307	DC0325	M67	M65	R	N	M56						NODCM
308	DR43490	M67	M65	R	N	J						STOP

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
309	DC0257	N	M70									STOP
310	DC0537	N	M59									NODCM
311	DR43618	N	R	M78								NODCM
312	DC0112	U										DUPLICATION
313	DC0147	U										NODCM
314	DR43369	U										DELETION
315	DR43908	B	J	N	O	Q	V					NODCM
316	DOBS0225	B	J	N	O	Q	V					NODCM
317	DR43193	J	N	O	Q	V						STOP
318	DOBS0226	B	J	N	V							NODCM
319	DC0770	B	C	N	O	Q	U					SPLICE
320	DOBS062	B	C	M42	N	U						NODCM
321	DC1058	B	K	N	O	S	V					NODCM
322	DC0235	M13	J	N	O	S	V					STOP
323	DOBS025	M13	M33	J	N	O	S	V				NODCM
324	OSU352	M13	N	O	S	V						DELETION
325	DC1218	M13	J	N	O	S	V					SUBEXONIC
326	DOBS0169	B	M26	M29	J	K	N	O	S	V		NODCM
327	DOBS042	B	M29	J	K	M47	N	O	S	V		NODCM
328	DOBS0214	M1	B	M29	J	K	M47	N	O	S	V	NODCM
329	DR43223	M29	J	K	M47	N	O	S	V			STOP
330	DRVH1040	F	J	N	O							DUPLICATION
331	DOBS0138	F										NODCM
332	DOBS0166	F										NODCM
333	DR0121	F										STOP
334	DR43427	F										SUBEXONIC
335	DRVH1009	J	N	R	M65	M67						STOP
336	DC0741	M1	J	N	R	M65	M67					NODCM
337	40818	D										NODCM
338	43033	D										SUBEXONIC
339	DC0011	D										NODCM
340	DC0019	D										STOP
341	DC0045	D										STOP
342	DC0189	D										NODCM
343	DC0277	D										SPLICE
344	DC0616	D										STOP
345	DC0733	D										STOP
346	DC0753	D										SUBEXONIC
347	DOBS0114	D										NOTDMD
348	DOBS0136	D										NODCM
349	DR43166	D										NODCM
350	DR43176	D										SUBEXONIC
351	DR43181	D										MISSENSE
352	DR43765	D										NODCM
353	DR43806	D										SUBEXONIC
354	DR43854	D										DUPLICATION
355	DR43862	D										NODCM
356	DRVH1037	D										DUPLICATION
357	OSU363	D										STOP
358	DOBS0200	D										NODCM
359	DC1170	D										STOP
360	DC1166	D										STOP

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)										Mutation
361	DC0619	E										POTENTIAL
362	DOBS0105	E										NOTDMD
363	DC0295	E										NODCM
364	DC0296	E										SUBEXONIC
365	DOBS0121	E										NOTDMD
366	DOBS078	E										NODCM
367	DR43608	E										NODCM
368	DC1182	E										SPLICE
369	DR43286	E	M30									SPLICE
370	43040	G										STOP
371	DC1077	G										NODCM
372	DC1174	G										NODCM
373	DC1161	G										STOP
374	DOBS056	M15	G									NODCM
375	42985	H										STOP
376	DC0333	H										NODCM
377	DOBS0109	H										NOTDMD
378	DR0038	H										NODCM
379	DR43106	H										SUBEXONIC
380	DC1232	H										STOP
381	DR43170	P										STOP
382	DC0334	P										NODCM
383	DC0391	P										SUBEXONIC
384	DC0393	P										NODCM
385	DOBS0176	P										NODCM
386	DOBS0178	P										NODCM
387	OSU366	P										NODCM
388	35752	M10	M54									STOP
389	DR43712	M12										DUPLICATION
390	DC0877	M14										POTENTIAL
391	DOBS040	M15										NODCM
392	DOBS058	M16										NODCM
393	DR43689	M17										STOP
394	OSU360	M17										SUBEXONIC
395	DOBS0113	M18	M32	M40								NOTDMD
396	DC0027	M2										SUBEXONIC
397	DC1073	M23										STOP
398	DR43100	M23										SPLICE
399	OSU315	M23										SPLICE
400	DC1229	M23										STOP
401	42993	M24										DELETION
402	DC0553	M25										SUBEXONIC
403	DOBS037	M28										NODCM
404	DOBS098	M31										NOTDMD
405	42963	M4										NOTDMD
406	DC0585	M40	O	R	M64	M71						NODCM
407	DC1200	M43										NODCM
408	DR43291	M44										STOP
409	DR43834	M45										DUPLICATION
410	DOBS0179	M46										NODCM
411	DOBS072	M48										NODCM
412	DRVH1048	M49										OTHER

Table B.2 continued

#	Patient ID	Variants found in each patient (Variant Code*)									Mutation
413	DR43721	M52									STOP
414	DC0176	M54									SUBEXONIC
415	DR43899	M57									DUPLICATION
416	DC1168	M57									NODCM
417	DR0112	M6	M3								NODCM
418	DOBS013	M60									NODCM
419	DC0751	M61									DELETION
420	DOBS0117	M62									NOTDMD
421	DR0190	M63									STOP
422	DR0125	M66									MISSENSE
423	DOBS017	M69									NODCM
424	OSU362	M7	M33								STOP
425	DR43821	M72									SUBEXONIC
426	DRVH43499	M73									SUBEXONIC
427	DC0568	M77									SUBEXONIC
428	DC1050	M77									STOP
429	DC1130	M78									NODCM
430	DR43704	M8									SPLICE
431	DOBS0234	M8									NODCM
432	DC0107	M80									NODCM
433	DOBS060	M9									NODCM
434	43078	M10									DELETION
435	DRVH1106	M10									DELETION

Note: *Each variant in each patient is depicted in this graph using the 'Variant Code' assigned to each variant in Table B.1. The variants for each patient are sorted from left to right by 3' UTR position, where the leftmost variants are found earlier in the 3' UTR. Patients described as having no mutations in the coding region of the *DMD* gene (NODCM) are indicated.

Table B.3

DMD 3' UTR variants reported in public databases.

#	Variant	3' UTR Position	3' UTR Region	Variant Code	# of Hits DMD Patients	# of Hits Leiden Database	dbSNP #	1000 Genome MAF
1	c.*15A>G	15	Aa		-	-	rs149405184	-
2	c.*23_*35del	23	Aa		-	4	-	-
3	c.*24G>A	24	Aa		-	-	rs372284841	-
4	c.*38G>A	38	Aa	M8	4	3	rs16989352	0.0067
5	c.*39T>A	39	Aa		-	-	rs371730838	-
6	c.*64_*65insC	64	Aa	M9	2	-	-	-
7	c.*93_*94insT	93	Aa	M10	4	-	-	-
8	c.*95T>C	95	Aa		-	-	rs369622987	-
9	c.*98G>C	98	Aa		-	-	rs377175004	-
10	c.*131T>C	131	Aa		-	-	rs404401	-
11	c.*132A>C	132	Aa		-	-	rs75075462	-
12	c.*136G>C	136	Aa		-	-	rs372759340	-
13	c.*169A>C	169	Ab		-	1	-	-
14	c.*175C>A	175	Ab		-	-	rs418795	-
15	c.*205A>G	205	Ab		-	-	rs5927686	-
16	c.*266C>T	266	Ab	M11	1	-	-	-
17	c.*269C>A	269	Ab		-	-	rs418560	-
18	c.*428G>C	428	Ac	M12	1	-	-	-
19	c.*[467C>T;477T.C]	467	Ad		-	1	-	-
20	c.*467_*476del	467	Ad	M13	4	2	rs59588979	-
21	c.*474_*475insAC	474	Ad		-	-	rs63123561	-
22	c.*469_*476delCACACACA	469	Ad		-	-	rs68131847	-
23	c.*471_*476delCACACA	471	Ad	M14	1	-	-	-
24	c.*476_*477insACAT	476	Ad		-	-	rs63332860	-
25	c.*473C>T	473	Ad		-	1	rs112111580	-
26	c.*475_*476dupCA	475	Ad	M15	3	-	-	-
27	c.*477_*480dupTACA	477	Ad	C	221	110	rs35017372	-
28	c.*477_*484del	477	Ad	D	24	20	-	-
29	c.*477_*490del	477	Ad	E	9	4	rs199552087	0.0169
30	c.*477_*488del	477	Ad	F	5	2	-	-
31	c.*477_*480delTACA	477	Ad	M16	2	-	-	-
32	c.*477_*482delTACACA	477	Ad	M17	2	1	-	-
33	c.*477_*486del	477	Ad	M18	1	-	rs200467976	0.0345
34	c.*477T>C	477	Ad	M19	1	1	-	-
35	c.*477_*478delTA	477	Ad		-	-	rs66790196	-
36	c.*480_*481insTA	480	Ad	M20	2	-	-	-
37	c.*480_*481insTACACA	480	Ad	M21	1	-	-	-
38	c.*480_*481insTACA	480	Ad		-	-	rs3833413	0.198
39	c.*481C>T	481	Ad		-	-	rs112509940	-
40	c.*483_*492del	483	Ad	M22	1	-	-	-
41	c.*489C>T	489	Ad	M23	4	2	rs45549534	-
42	c.*489_*492dupCACA	489	Ad	M24	1	1	rs72287742, rs72300588	-

Table B.3 continued

#	Variant	3' UTR Position	3' UTR Region	Variant Code	# of Hits DMD Patients	# of Hits Leiden Database	dbSNP #	1000 Genome MAF
43	c.*491_*492delCA	491	Ad	G	5	2	-	-
44	c.*491_*492dupCA	491	Ad	M25	2	3	-	-
45	c.*507C>T	507	Ad		-	-	rs185632747	0.001
46	c.*571A>G	571	Ad	M26	1	-	rs192963364	0.005
47	c.*627A>G	627	Ad/Ma		-	-	rs187363115	0.002
48	c.*645G>A	645	Ad/Ma		-	-	rs182687817	0.001
49	c.*651G>A	651	Ma	M27	1	-	rs149425144	0.002
50	c.*686T>G	686	Ma		-	-	rs189948189	0.0012
51	c.*755_*758dupATTT	755	Ma	M28	1	-	-	-
52	c.*809C>T	809	Ma	M29	4	-	rs138956803	0.0115
53	c.*855G>A	855	Ma		-	-	rs150236690	0.0042
54	c.*921_*924dupACAA	921	Ma	M30	1	2	rs72466531	-
55	c.*952_*956delAAGTA	952	Ma	M31	2	80	-	-
56	c.*952_*953dupAA	952	Ma	M32	1	-	rs60841052, rs367926291, rs200963373	0.0073
57	c.*955_956delTA	955	Ma			-	rs368966903	-
58	c.*968_*969insT	968	Ma	M33	3	3	rs72466530	-
59	c.*998_*1001dupCTTA	998	Ma	M34	2	-	-	-
60	c.*1006_*1009dupATCA	1006	Ma	M35	1	-	-	-
61	c.*1015A>C	1015	Ma/Mb		-	-	rs186667089	0.0006
62	c.*1022_*1023insATTG	1022	Ma/Mb	M36	1	-	-	-
63	c.*1023G>A	1023	Mb		-	-	rs111354150	0.0024
64	c.*1031_*1038del	1031	Mb	M37	4	-	rs201552499	0.0206
65	c.*1031_*1042del	1031	Mb	M38	1	-	-	-
66	c.*1031G>A	1031	Mb	M39	1	-	rs112666076	0.02
67	c.*1030_*1031insTTGA	1030	Mb		-	-	rs62871360	-
68	c.*1031_*1034delGTTG	1031	Mb		-	-	rs200048960	0.0617
69	c.*1035A>G	1035	Mb		-	-	rs112754560	-
70	c.*1045_*1052dupTGATTGAT	1045	Mb	M40	3	-	-	-
71	c.*1045delT	1045	Mb		-	-	rs202075326	0.0169
72	c.*1046_*1048dupGAT	1046	Mb	H	6	4	rs71860126	0.1252
73	c.*1048delT	1048	Mb		-	1	rs72466529	-
74	c.*1049_*1052dupTGAT	1049	Mb	I	206	186	rs72082375, rs111646375, rs59846267	-
75	c.*1049_*1052delTGAT	1049	Mb	J	18	10	rs200641250	-
76	c.*1052_*1053insTGGTTGAT TGATTGATTGATTGAT	1052	Mb	M41	1	-	-	-
77	c.*1052_*1053insTGGTTGAT TGATTGATTGATTGATTGAT	1052	Mb	M42	1	-	-	-
78	c.*1165C>T	1165	Mb	M43	1	-	-	-
79	c.*1182C>T	1182	Mb		-	-	rs187318961	0.0073
80	c.*1185_*1188dupTGAT	1185	Mb	K	5	-	rs200311985	0.0091
81	c.*1230G>C	1230	Mb		-	-	rs145871126	0.0006
82	c.*1242_*1245dupCTTT	1242	Mb	M44	1	2	rs72466527	-

Table B.3 continued

#	Variant	3' UTR Position	3' UTR Region	Variant Code	# of Hits DMD Patients	# of Hits Leiden Database	dbSNP #	1000 Genome MAF
83	c.*1285C>A	1285	Mb	L	6	-	rs142110702	0.0139
84	c.*1353_*1354insT	1353	Mb	M45	1	-	-	-
85	c.*1393T>A	1393	Mb	M46	1	-	-	-
86	c.*1401C>A	1401	Mb	M47	3	-	rs182383454	0.0024
87	c.*1447A>G	1447	Mb	N	321	176	rs3361	0.3337
88	c.*1455C>A	1455	Mb		-	-	rs191747923	0.0018
89	c.*1476G>A	1476	Mb	M48	1	-	-	-
90	c.*1533G>A	1533	Mc	M49	1	-	-	-
91	c.*1562_*1565dupTAAG	1562	Mc	O	70	2	rs372204167, rs202212782	0.078
92	c.*1565insAAA	1565	Mc		-	-	rs72466526	-
93	c.*1588_*1591dupTGTT	1588	Mc	P	7	5	rs72466525	-
94	c.*1600_*1601insA	1600	Mc	M50	1	-	-	-
95	c.*1638A>G	1638	Mc	Q	5	-	rs142520583	0.0115
96	c.*1654C>G	1654	Mc		-	-	rs148472346	0.0018
97	c.*1662insA	1662	Mc		-	-	rs373359794	-
98	c.*1705A>G	1705	Mc	M51	1	-	rs145632098	0.0036
99	c.*1716_*1717insC	1716	Mc	M52	1	-	-	-
100	c.*1731A>C	1731	Mc	M53	1	-	rs3198427	0.0042
101	c.*1750_*1751insA	1750	Mc	M54	1	-	-	-
102	c.*1758_*1761delTTCA	1758	Mc	M55	4	6	-	-
103	c.*1758_*1759insTCAT	1758	Mc		-	-	rs72466524	-
104	c.*1759delT	1759	Mc		-	1	-	-
105	c.*1762delA	1762	Mc		-	1	-	-
106	c.*1783A>G	1783	Mc	R	9	7	rs7886658	0.0496
107	c.*1799G>T	1799	Mc	S	9	-	rs16989350	0.0369
108	c.*1826C>G	1826	Mc		-	-	rs112358696	-
109	c.*1838_*1841delTGAT	1838	Mc	M56	1	-	-	-
110	c.*1903_*1906dupTTAA	1903	Mc/Md	M57	2	-	-	-
111	c.*1903_*1906delTTAA	1903	Mc/Md		-	3	-	-
112	c.*1947G>A	1947	Mc/Md		-	-	rs185921823	-
113	c.*1951_*1952insT	1951	Md	M58	2	-	rs201927678	0.0085
114	c.*2018_*2021dupCTTT	2018	Md	M59	1	-	-	-
115	c.*2019T>C	2019	Md	M60	1	-	-	-
116	c.*2028_*2048del	2028	Md	M61	1	-	-	-
117	c.*2030_*2033dupTAAT	2030	Md	M62	1	-	-	-
118	c.*2031_*2034dupAATT	2031	Md	M63	1	-	-	-
119	c.*2037_*2038dupTT	2037	Md	M64	1	-	rs200703281	0.0097
120	c.*2039C>A	2039	Md	M65	4	5	rs72466523	0.0266
121	c.*2103T>A	2103	Md	M66	1	-	-	-
122	c.*2139_*2155dup	2139	Md		-	1	-	-
123	c.*2153_*2154insATGTGACGCTGG	2153	Md		-	4	-	-
124	c.*2155_*2156insATGTGACGCTGGACCTT	2155	Md	M67	4	-	rs200762890, rs72466522	0.0151
125	c.*2160_*2161insTCTTT	2160	Md		-	-	rs369593541	-

Table B.3 continued

#	Variant	3' UTR Position	3' UTR Region	Variant Code	# of Hits DMD Patients	# of Hits Leiden Database	dbSNP #	1000 Genome MAF
126	c.*2182_*2183insCTCA	2182	Md		-	-	rs375086194	-
127	c.*2185A>T	2185	Md		-	-	rs11550191	-
128	c.*2234_*2237dupAAGT	2234	Da	T	34	12	rs67121194, rs72466521, rs202154420, rs199643329	0.0181
129	c.*2245T>C	2245	Da	M68	1	-	-	-
130	c.*2252A>G	2252	Da		-	3	-	-
131	c.*2263_*2264insATAGA	2263	Da		-	-	rs369375730	-
132	c.*2268C>T	2268	Da	M69	1	-	-	-
133	c.*2314T>C	2314	Da	M70	1	-	-	-
134	c.*2326G>A	2326	Da	M71	1	-	-	-
135	c.*2342T>C	2342	Da		-	-	rs181280702	-
136	c.*2427A>G	2427	Db		-	-	rs1057915	-
137	c.*2548_*2549insC	2548	Db	M72	1	-	-	-
138	c.*2563G>A	2563	Db/Dc		-	-	rs188558013	0.0024
139	c.*2578_*2581dupACTT	2578	Db/Dc	M73	1	2	rs72466520	-
140	c.*2591T>A	2591	Dc	M74	1	-	-	-

Note: *DMD* variants found in all databases searched are shown. Variant codes were assigned for all variants found in the DMD patient population (See Table B.1). The number of patients found in the Leiden Database (Aartsma-Rus et al., 2006) is shown (# of Hits Leiden Database). dbSNP numbers are shown for any variant found in the current version of dbSNP (Sherry et al., 2001). In some cases, multiple dbSNP numbers describe the same mutation. For example, a four base pair duplication may be described as a duplication or insertion and given two separate dbSNP numbers. For any variant reported in the 1000 genomes project (Abecasis et al., 2012), the Minor Allele Frequency (MAF) is shown. A '-' indicates that the variant was not found in the database.

Table B.4

Predicted miRNA binding sites in the *DMD* 3' UTR.

miRNA	3' UTR Position		Binding Sequence	3' UTR region	Comments
	Start	End			
miR-30a-e	158	164	GUUUACA	Ab	Downregulated in human DMD patients [1]; Downregulated in *mdx* mice [2]; Involved in muscle differentiation during zebrafish embryonic development [3]
miR-144	203	210	AUACUGUA	Ab	The variant c.*205A>G, reported in dbSNP (rs5927686)
miR-101	204	210	UACUGUA	Ab	The variant c.*205A>G, reported in dbSNP (rs5927686); Downregulated in human DMD patients [1]
miR-139-5p	205	212	ACUGUAGA	Ab	The variant c.*205A>G, reported in dbSNP (rs5927686); Downregulated in *mdx* mice [2]
miR-153	269	276	CUAUGCAA	Ab	The variant c.*269C>A, reported in dbSNP (rs418560)
miR-129-5p	282	288	CAAAAAA	Ab	
miR-31	330	336	CUUGCCA	Ac	Upregulated in damaged myofibers [4]; Regulates dystrophin expression [5]
let-7a-f; miR-98; miR-4458; miR-4500	1934	1941	CUACCUCA	Mc/Md	
miR-190	2603	2610	ACAUAUCA	Dc	
miR-124; miR-506	2627	2633	UGCCUUA	Dc	

Note: Predicted miRNA binding sites in the *DMD* 3' UTR using TargetScan (Lewis et al., 2005) are shown. The 3' UTR regions deleted that contain each predicted binding site are shown (see Chapter 1 for details of deleted regions). References: [1] - (Eisenberg et al., 2007); [2] - (Roberts et al., 2012); [3] - (Ketley et al., 2013); [4] - (Greco et al., 2009); [5] - (Cacchiarelli et al., 2011).

Table B.5

Predicted RNA protein binding sites in the Lemaire A region.

#	Score	Relative Score	RBP Name	Start	End	Matching sequence	3' UTR Region	Human variants within or adjacent to predicted binding sites
1	4.67	88%	RBMX	11	14	CCAC	Aa1	c.*15A>G, reported in dbSNP (rs149405184)
2	3.83	86%	ELAVL1 (HuR)	25	28	AUUU	Aa1	c.*23_*35del, reported in Leiden Database (4 patients)
3	4.19	90%	SFRS1	40	43	UGGA	Aa2	
4	5.01	80%	YTHDC1	42	47	GAGUCC	Aa2	
5	6.34	100%	KHSRP	44	47	GUCC	Aa2	
6	6.63	93%	SFRS9	74	78	AGGAG	Aa3	
7	4.62	100%	SFRS1	74	77	AGGA	Aa3	
8	4.33	92%	KHDRBS3	82	87	AAUAAA	Aa3	
9	4.40	100%	ELAVL1 (HuR)	89	92	GUUU	Aa3	c.*93_*94insT, found in DMD population (M9 in 2 patients)
10	4.40	83%	RBMX	107	110	CCCG	Aa3	
11	8.65	100%	a2bp1	110	114	GCAUG	Aa3/Aa4	
12	4.40	100%	ELAVL1 (HuR)	115	118	GUUU	Aa4	
13	4.60	98%	KHDRBS3	120	125	UAUAAU	Aa4	
14	5.63	90%	YTHDC1	128	133	UCAUAC	Aa4	c.*131T>C, reported in dbSNP (rs404401); c.*132A>C, reported in dbSNP (rs75075462)
15	4.77	93%	SFRS13A	137	143	AAAGAGG	Aa4	c.*136G>C, reported in dbSNP (rs372759340)
16	4.14	81%	SFRS13A	139	145	AGAGGAU	Aa4	
17	4.62	100%	SFRS1	141	144	AGGA	Aa4	
18	4.40	100%	ELAVL1 (HuR)	158	161	GUUU	Ab	
19	4.33	92%	KHDRBS3	168	173	AAUAAA	Ab	c.*169A>C, reported in Leiden Database (1 patient)
20	3.83	86%	ELAVL1 (HuR)	179	182	AUUU	Ab	
21	7.56	86%	sap-49	183	188	UUGUGA	Ab	
22	7.23	100%	Pum2	207	210	UGUA	Ab	
23	3.83	86%	ELAVL1 (HuR)	212	215	AUUU	Ab	
24	4.40	100%	ELAVL1 (HuR)	221	224	GUUU	Ab	
25	4.40	100%	ELAVL1 (HuR)	239	242	GUUU	Ab	
26	4.40	100%	ELAVL1 (HuR)	258	261	GUUU	Ab	
27	7.23	100%	Pum2	278	281	UGUA	Ab	
28	8.72	100%	PABPC1	283	287	AAAAA	Ab	
29	8.72	100%	PABPC1	284	288	AAAAA	Ab	
30	7.23	100%	Pum2	306	309	UGUA	Ac	
31	3.83	82%	KHDRBS3	306	311	UGUAAA	Ac	
32	3.94	84%	KHDRBS3	321	326	GCUAAA	Ac	
33	4.02	86%	KHDRBS3	325	330	AAUAAC	Ac	
34	5.00	94%	RBMX	334	337	CCAU	Ac	
35	3.83	86%	ELAVL1 (HuR)	336	339	AUUU	Ac	
36	4.19	90%	SFRS1	347	350	UGGA	Ac	
37	6.47	100%	EIF4B	348	351	GGAA	Ac	
38	3.83	86%	ELAVL1 (HuR)	355	358	AUUU	Ac	
39	10.35	80%	KHDRBS3	363	378	UUGUUUAAAAA UUUAU	Ac	
40	4.40	100%	ELAVL1 (HuR)	365	368	GUUU	Ac	
41	4.04	86%	KHDRBS3	366	371	UUUAAA	Ac	
42	8.72	100%	PABPC1	369	373	AAAAA	Ac	
43	3.83	86%	ELAVL1 (HuR)	373	376	AUUU	Ac	
44	3.97	85%	KHDRBS3	376	381	UAUAAC	Ac	
45	4.28	91%	KHDRBS3	385	390	UAUAAA	Ac	
46	4.53	88%	SFRS13A	388	394	AAAGAAA	Ac	
47	4.32	84%	SFRS13A	392	398	AAAGAUU	Ac	
48	7.23	100%	Pum2	398	401	UGUA	Ac	
49	3.83	82%	KHDRBS3	398	403	UGUAAA	Ac	
50	4.14	89%	KHDRBS3	403	408	ACUAAA	Ac	

Table B.5 continued

#	Score	Relative Score	RBP Name	Start	End	Matching sequence	3' UTR Region	Human variants within or adjacent to predicted binding sites
51	6.63	100%	MBNL1	412	415	UGCU	Ac	
52	4.28	91%	KHDRBS3	417	422	UAUAAA	Ac	
53	8.72	100%	PABPC1	420	424	AAAAA	Ac	
54	8.72	100%	PABPC1	421	425	AAAAA	Ac	
55	8.72	100%	PABPC1	422	426	AAAAA	Ac	
56	8.72	100%	PABPC1	423	427	AAAAA	Ac	c.*428G>C, found in DMD patients (Variant code M12 in 1 patient)
57	4.40	100%	ELAVL1 (HuR)	431	434	GUUU	Ac/Ad	
58	4.28	91%	KHDRBS3	434	439	UAUAAA	Ac/Ad	
59	8.72	100%	PABPC1	437	441	AAAAA	Ac/Ad	
60	4.09	88%	KHDRBS3	444	449	CCUAAA	Ad	
61	8.72	100%	PABPC1	447	451	AAAAA	Ad	
62	7.15	84%	PABPC1	456	462	ACAAACA	Ad	
63	7.66	88%	RBMY1A1	489	493	CACAA	Ad	c.*489C>T, found in DMD patients (M23 in 4 patients), reported in Leiden database (2 patients), reported in dbSNP (rs45549534); c.*489_*492dupCACA, found in DMD patients (M24 in 1 patient), reported in Leiden database (1 patient), reported in dbSNP (rs72287742, rs72300588)
64	4.40	100%	ELAVL1 (HuR)	513	516	GUUU	Ad	
65	5.18	83%	YTHDC1	518	523	GCAUCC	Ad	
66	7.56	86%	sap-49	529	534	GCGUGA	Ad	
67	5.00	94%	RBMX	538	541	CCAU	Ad	
68	4.09	87%	KHDRBS3	573	578	AUUAAA	Ad	
69	4.11	80%	SFRS13A	576	582	AAAGAUA	Ad	
70	4.67	88%	RBMX	595	598	CCAC	Ad	
71	4.67	88%	RBMX	598	601	CCAC	Ad	
72	5.28	84%	YTHDC1	609	614	GACUAC	Ad	
73	5.02	80%	YTHDC1	612	617	UACUAC	Ad	
74	6.63	100%	MBNL1	622	625	UGCU	Ad	

Note: Predicted RNA protein binding sites in the Lemaire A region of the *DMD* 3' UTR using RBPDB (Cook et al., 2011) are shown. RBPDB predicts binding sites by scoring potential binding sites in the sequence using position weight matrices and assigns a score that is calculated as the sum of the scores of each nucleotide at each position in the position weight matrix. A relative score is also given as a percent of the score to the maximum possible score of the calculated matrix. Only predicted sites with a relative score above 80% are shown. The 3' UTR start and end positions if each predicted binding site are shown. Human variants found within or directly adjacent to a predicted binding site are noted.

Table B.6

Predicted RNA protein binding sites in a conserved AU rich region.

#	Score	Relative Score	RBP Name	Start	End	Matching sequence	3' UTR Region	Human variants within or adjacent to predicted binding sites
1	8.785	100%	sap-49	1786	1791	GUGUGA	Mc	
2	4.604	98%	KHDRBS3	1794	1799	UAUAAU	Mc	c.*1799G>T, found in DMD patients (S
3	3.826	86%	ELAVL1 (HuR)	1798	1801	AUUU	Mc	in 9 patients, reported in dbSNP
4	7.229	100%	Pum2	1829	1832	UGUA	Mc	
5	4.153	89%	KHDRBS3	1829	1834	UGUAAU	Mc	
6	4.188	90%	SFRS1	1842	1845	UGGA	Mc	
7	7.229	100%	Pum2	1864	1867	UGUA	Mc/Md	
8	4.153	89%	KHDRBS3	1864	1869	UGUAAU	Mc/Md	
9	4.188	90%	SFRS1	1877	1880	UGGA	Mc/Md	
10	4.404	100%	ELAVL1 (HuR)	1887	1890	GUUU	Mc/Md	
11	6.34	100%	KHSRP	1892	1895	GUCC	Mc/Md	
12	4.999	94%	RBMX	1894	1897	CCAU	Mc/Md	
13	4.419	94%	KHDRBS3	1899	1904	AUUAAU	Mc/Md	c.*1903_*1906dupTTAA, found in
14	4.652	100%	KHDRBS3	1902	1907	AAUAAU	Mc/Md	DMD patients (M57 in 2 patients);
15	4.419	94%	KHDRBS3	1906	1911	AUUAAU	Mc/Md	
16	4.419	94%	KHDRBS3	1910	1915	AUUAAU	Mc/Md	
17	5.739	89%	QKI	1913	1922	AAUUAACAUC	Mc/Md	
18	3.788	81%	KHDRBS3	1914	1919	AUUAAC	Mc/Md	
19	9.428	100%	ybx2-a	1917	1922	AACAUC	Mc/Md	
20	9.374	100%	ybx2-a	1917	1922	AACAUC	Mc/Md	
21	5.629	90%	YTHDC1	1935	1940	UCAUGC	Mc/Md	
22	6.628	100%	MBNL1	1938	1941	UGCU	Mc/Md	
23	3.826	86%	ELAVL1 (HuR)	1942	1945	AUUU	Mc/Md	
24	4.404	100%	ELAVL1 (HuR)	1959	1962	GUUU	Mc/Md	
25	7.369	100%	FUS	1966	1969	GGUG	Md	
26	4.449	95%	KHDRBS3	1975	1980	GAUAAU	Md	
27	4.667	88%	RBMX	2007	2010	CCAC	Md	
28	6.34	100%	KHSRP	2013	2016	GUCC	Md	
29	4.999	94%	RBMX	2015	2018	CCAU	Md	c.*2018_*2021dupCTTT, found in DMD patients (M59 in 1 patient); c.*2019T>C, found in DMD Patients (M60 in 1 patient)
30	3.826	86%	ELAVL1 (HuR)	2025	2028	AUUU	Md	c.*2028_*2048del, found in DMD patients (M61 in 1 patient)
31	4.604	98%	KHDRBS3	2040	2045	UAUAAU	Md	c.*2039C>A, found in DMD patients (M65 in 4 patients), reported in Leiden database (5 patients), reported in dbSNP (rs72466523), reported in 1000 genome project (MAF = 0.0266)
32	4.604	98%	KHDRBS3	2061	2066	CAUAAU	Md	
33	3.826	86%	ELAVL1 (HuR)	2065	2068	AUUU	Md	
34	4.529	88%	SFRS13A	2072	2078	AAAGAAA	Md	

Note: Predicted RNA protein binding sites contained within a conserved AU rich sequence found in the middle region of the *DMD* 3' UTR using RBPDB (Cook et al., 2011) are shown. RBPDB predicts binding sites by scoring potential binding sites in the sequence using position weight matrices and assigns a score that is calculated as the sum of the scores of each nucleotide at each position in the position weight matrix. A relative score is also given as a percent of the score to the maximum possible score of the calculated matrix. Only predicted sites with a relative score above 80% are shown. The 3' UTR start and end positions of each predicted binding site are shown. Human variants found within or directly adjacent to a predicted binding site are noted.

Table B.7

Predicted RNA protein binding sites in the Lemaire D region.

#	Score	Relative Score	RBP Name	Start	End	Matching sequence	3' UTR Region	Human variants within or adjacent to predicted binding sites
1	4.404	100%	ELAVL1 (HuR)	2248	2251	GUUU	Da	c.*2252A>G, reported in Leiden database (3 patients)
2	4.094	88%	KHDRBS3	2255	2260	UCUAAA	Da	
3	10.3	100%	ZRANB2	2266	2271	AGGUAA	Da	c.*2268C>T, reported in DMD patients (M69 in 1 patient)
4	5.715	91%	YTHDC1	2275	2280	GAGUGC	Da	
5	4.277	91%	KHDRBS3	2280	2285	CAUAAA	Da	
6	4.652	100%	KHDRBS3	2284	2289	AAUAAU	Da	
7	3.826	86%	ELAVL1 (HuR)	2288	2291	AUUU	Da	
8	11.44	81%	ELAVL2	2289	2297	UUUUGUUUU	Da	
9	10.99	85%	ELAVL2	2289	2297	UUUUGUUUU	Da	
10	4.404	100%	ELAVL1 (HuR)	2293	2296	GUUU	Da	
11	4.371	93%	KHDRBS3	2295	2300	UUUAAU	Da	
12	11.44	81%	ELAVL2	2303	2311	UUUUGUUUU	Da	
13	10.99	85%	ELAVL2	2303	2311	UUUUGUUUU	Da	
14	4.404	100%	ELAVL1 (HuR)	2307	2310	GUUU	Da	
15	4.188	90%	SFRS1	2332	2335	UGGA	Da	
16	4.604	98%	KHDRBS3	2346	2351	CAUAAU	Da	
17	3.826	86%	ELAVL1 (HuR)	2352	2355	AUUU	Da	
18	8.718	100%	PABPC1	2360	2364	AAAAA	Da	
19	6.628	100%	MBNL1	2378	2381	UGCU	Da	
20	3.826	86%	ELAVL1 (HuR)	2384	2387	AUUU	Da	
21	4.604	98%	KHDRBS3	2392	2397	CAUAAU	Da	
22	4.419	94%	KHDRBS3	2396	2401	AUUAAU	Da	
23	3.826	86%	ELAVL1 (HuR)	2400	2403	AUUU	Da	
24	4.188	90%	SFRS1	2403	2406	UGGA	Da	
25	5.257	81%	EIF4B	2404	2407	GGAC	Da	
26	3.826	86%	ELAVL1 (HuR)	2411	2414	AUUU	Db	
27	3.973	85%	KHDRBS3	2429	2434	UAUAAC	Db	
28	4.667	88%	RBMX	2434	2437	CCAC	Db	
29	4.092	87%	KHDRBS3	2443	2448	AUUAAA	Db	
30	7.229	100%	Pum2	2450	2453	UGUA	Db	
31	3.826	82%	KHDRBS3	2450	2455	UGUAAA	Db	
32	4.604	98%	KHDRBS3	2457	2462	CAUAAU	Db	
33	7.229	100%	Pum2	2462	2465	UGUA	Db	
34	4.277	91%	KHDRBS3	2473	2478	CAUAAA	Db	
35	9.428	100%	ybx2-a	2477	2482	AACAUC	Db	
36	9.374	100%	ybx2-a	2477	2482	AACAUC	Db	
37	8.647	100%	a2bp1	2488	2492	GCAUG	Db	
38	4.404	100%	ELAVL1 (HuR)	2492	2495	GUUU	Db	
39	4.404	100%	ELAVL1 (HuR)	2503	2506	GUUU	Db	
40	4.604	98%	KHDRBS3	2534	2539	CAUAAU	Db	
41	4.404	100%	ELAVL1 (HuR)	2547	2550	GUUU	Db	c.*2548_*2549insC, found in DMD patients (M72 in 1 patient)
42	5.685	89%	QKI	2548	2557	UUUUAACACC	Db	
43	3.739	80%	KHDRBS3	2549	2554	UUUAAC	Db/Dc	
44	7.893	84%	ybx2-a	2552	2557	AACACC	Db/Dc	
45	7.229	100%	Pum2	2563	2566	UGUA	Db/Dc	c.*2563G>A, reported in dbSNP (rs188558013), reported in 1000 genome project (MAF = 0.0024)
46	3.826	86%	ELAVL1 (HuR)	2569	2572	AUUU	Db/Dc	
47	3.826	86%	ELAVL1 (HuR)	2580	2583	AUUU	Dc	
48	4.044	86%	KHDRBS3	2585	2590	UUUAAA	Dc	c.*2591T>A, reported in DMD patients (M74 in 1 patient)

Table B.7 continued

#	Score	Relative Score	RBP Name	Start	End	Matching sequence	3' UTR Region	Human variants within or adjacent to predicted binding sites
49	4.404	100%	ELAVL1 (HuR)	2596	2599	GUUU	Dc	
50	5.024	80%	YTHDC1	2600	2605	UACUGC	Dc	
51	3.826	86%	ELAVL1 (HuR)	2606	2609	AUUU	Dc	
52	7.656	88%	RBMY1A1	2611	2615	CACAA	Dc	
53	7.229	100%	Pum2	2650	2653	UGUA	Dc	
54	5.46	87%	YTHDC1	2657	2662	UAGUAC	Dc	
55	5.024	80%	YTHDC1	2660	2665	UACUGC	Dc	

Note: Predicted RNA protein binding sites in the Lemaire D region of the *DMD* 3' UTR using RBPDB (Cook et al., 2011) are shown. RBPDB predicts binding sites by scoring potential binding sites in the sequence using position weight matrices and assigns a score that is calculated as the sum of the scores of each nucleotide at each position in the position weight matrix. A relative score is also given as a percent of the score to the maximum possible score of the calculated matrix. Only predicted sites with a relative score above 80% are shown. The 3' UTR start and end positions if each predicted binding site are shown. Human variants found within or directly adjacent to a predicted binding site are noted.

References

Aartsma-Rus, A., Van Deutekom, J.C., Fokkema, I.F., Van Ommen, G.J., and Den Dunnen, J.T. (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve *34*, 135-144.

Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

Cacchiarelli, D., Incitti, T., Martone, J., Cesana, M., Cazzella, V., Santini, T., Sthandier, O., and Bozzoni, I. (2011). miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy. EMBO Rep *12*, 136-141.

Cook, K.B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T.R. (2011). RBPDB: a database of RNA-binding specificities. Nucleic Acids Res *39*, D301-308.

den Dunnen, J.T., and Antonarakis, S.E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat *15*, 7-12.

Eisenberg, I., Eran, A., Nishino, I., Moggio, M., Lamperti, C., Amato, A.A., Lidov, H.G., Kang, P.B., North, K.N., Mitrani-Rosenbaum, S.*, et al.* (2007). Distinctive patterns of microRNA expression in primary muscular disorders. Proc Natl Acad Sci U S A *104*, 17016-17021.

Greco, S., De Simone, M., Colussi, C., Zaccagnini, G., Fasanaro, P., Pescatori, M., Cardani, R., Perbellini, R., Isaia, E., Sale, P.*, et al.* (2009). Common micro-RNA signature in skeletal muscle damage and regeneration induced by Duchenne muscular dystrophy and acute ischemia. FASEB J *23*, 3335-3346.

Ketley, A., Warren, A., Holmes, E., Gering, M., Aboobaker, A.A., and Brook, J.D. (2013). The miR-30 microRNA family targets smoothened to regulate hedgehog signalling in zebrafish early muscle development. PLoS One *8*, e65170.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Roberts, T.C., Blomberg, K.E., McClorey, G., El Andaloussi, S., Godfrey, C., Betts, C., Coursindel, T., Gait, M.J., Smith, C.I., and Wood, M.J. (2012). Expression analysis in multiple muscle groups and serum reveals complexity in the microRNA transcriptome of the mdx mouse with implications for therapy. Mol Ther Nucleic Acids *1*, e39.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and

Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res *29*, 308-311.

APPENDIX C

CONTROLS FOR THE *DMD* 3' UTR EXPERIMENTS

Expression of the pHRL and 3' UTR constructs during C2C12 differentiation

Previously, we showed that a Renilla construct containing the *DMD* 3' UTR increased in expression during C2C12 differentiation relative to the control pHRL vector (Chapter 1). To determine whether the relative increase was due to an increase in the *DMD* 3' UTR construct or a decrease in the control pHRL construct during differentiation, we transfected C2C12 cells with these constructs and monitored changes in expression of the pHRL-CMV (pHRL) control construct and the *DMD* 3' UTR (3' UTR) construct every 24 hours for 7 days as the cells differentiated (Figure C.1). The observed increase in expression of the 3' UTR construct (3' UTR) began 2 days after C2C12 cells were grown in differentiation media and continued to increase until the 5th day of differentiation (Figure C.1). This increase in expression of the 3' UTR construct correlates with the morphology of the differentiating C2C12 cells with elongation of the myoblasts first being observed at Day 2 and large, mature myotubes being formed by Day 5. Expression of the control pHRL construct (pHRL) remained consistent during differentiation (Figure C.1). Construct synthesis, C2C12 transfections, and measurements of expression using a Renilla/Firefly dual luciferase assay were done as previously described (Chapter 1).

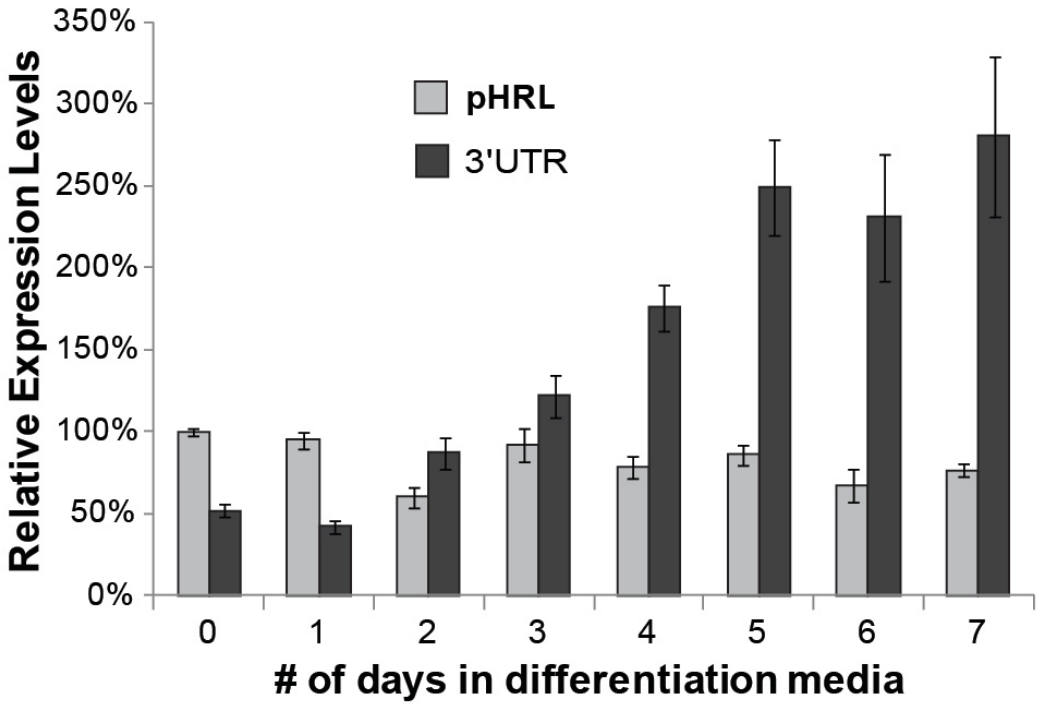


Figure C.1. Expression of the control pHRL construct remains consistent during C2C12 myogenesis.

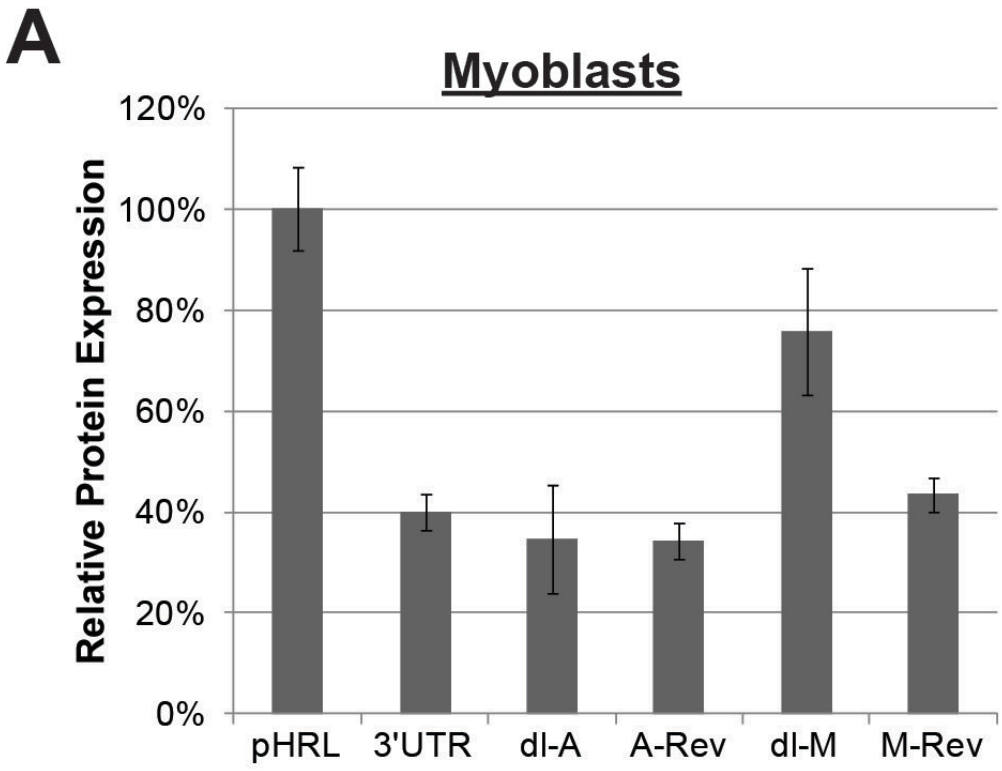
Relative expression levels for the control pHRL reporter construct (pHRL) and the *DMD* 3'UTR construct (3'UTR) are shown during the course of differentiation. C2C12 cells transfected with the *DMD* 3' UTR construct (3' UTR) had an increase in Renilla expression as the cells differentiated, whereas cells transfected with the control pHRL-CMV construct (pHRL) did not show an increase during differentiation. The average of four biological transfection replicates is shown for each construct at each time point. Error bars equal +/- 1 standard deviation.

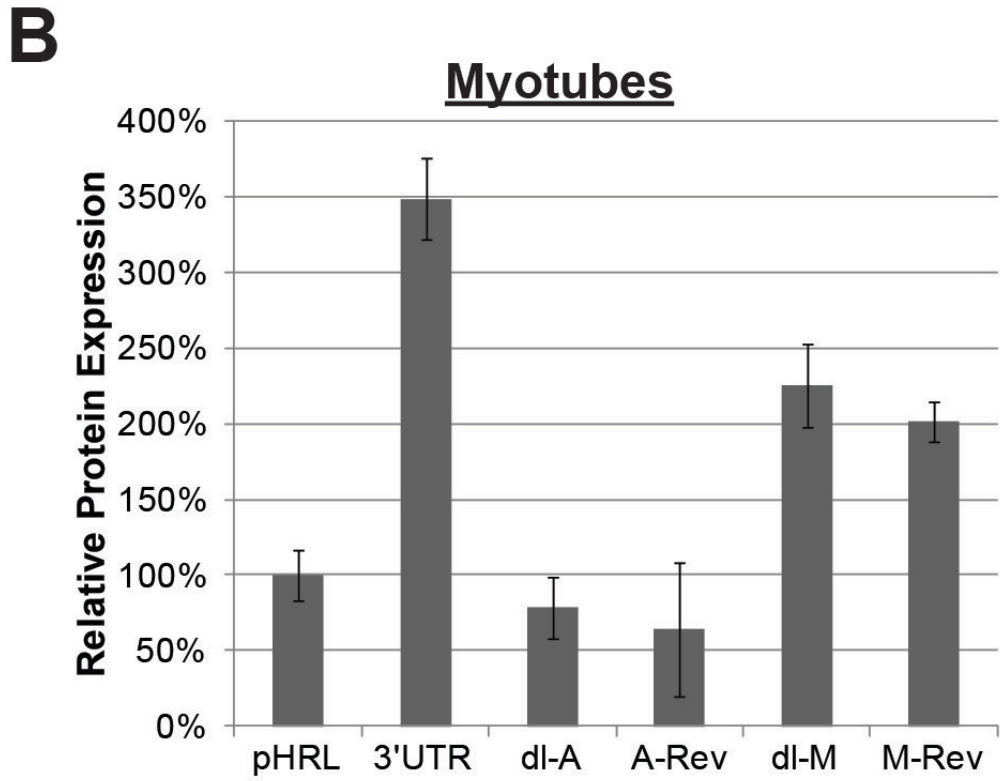
Size control constructs for deletions of the *DMD* 3' UTR

There is evidence that length of a 3' UTR can affect translational efficiency and mRNA stability (Hogg and Goff, 2010; Tanguay and Gallie, 1996). To determine if the effect of deleting regions of the 3' UTR was due to changing the size of the 3' UTR rather than deleting regulatory elements, we made *DMD* 3' UTR Renilla constructs containing the full length 3' UTR except with the Lemaire A or middle region inserted in the reverse orientation (A-rev and M-Rev, respectively), and transfected these constructs into C2C12 cells. Reversing the sequence would disrupt any regulatory elements contained within this region without changing the length of the 3' UTR. In C2C12 myoblasts and myotubes, deleting the Lemaire A region (dl-A) had the same effect on expression as reversing the Lemaire A region (A-Rev) (Figure C.2), showing that the decrease in expression in myotubes when this region is deleted is not due to changing the length of the 3'UTR. Likewise, deleting the middle region (dl-M) compared to reversing the middle region (M-Rev) resulted in the same level of expression in C2C12 myotubes (Figure C.2, Myotubes). However, deleting the middle region (dl-M) increased expression ~80% whereas reversing the middle region (M-Rev) did not increase expression compared to the full length 3' UTR construct in C2C12 myoblasts (Figure C.2, Myoblasts). The change in size of the 3' UTR when the middle region of the 3' UTR is deleted could account for the increase in expression seen in myoblasts. Alternatively, the previously described antisense lncRNA that spans the entire middle region of the *DMD* 3' UTR (Bovolenta et al., 2012) could still be functional when the middle region is reversed. The beginning and end portions of the lncRNA are found in the conserved regions of the 3' UTR and not deleted in our dl-M construct. Perhaps the

Figure C.2. Size control constructs for the dl-A and dl-M constructs in C2C12 cells.

A. Relative protein expression of the control Renilla vector (pHRL), the full length *DMD* 3' UTR construct (3'UTR), the 3' UTR constructs with the Lemaire A region or middle region deleted (dl-A and dl-M), and the 3' UTR constructs with the Lemaire A region or middle region reversed (A-Rev and M-Rev) in transfected C2C12 myoblasts are shown. B. Relative protein expression of the pHRL, 3' UTR and size control constructs transfected into C2C12 myotubes are shown. The result of three biological replicates is shown with expression levels normalized to pHRL-CMV expression (pHRL) using a Renilla/Firefly dual luciferase assay. Error bars equal +/- 1 standard deviation.


A
Myoblasts
Relative Protein Expression
120%
100%
80%
60%
40%
20%
0%
pHRL
3'UTR
dl-A
A-Rev
dl-M
M-Rev


B
Myotubes
Relative Protein Expression
400%
350%
300%
250%
200%
150%
100%
50%
0%
pHRL
3'UTR
dl-A
A-Rev
dl-M
M-Rev

lncRNA is still expressed and needs only to complement the 3' UTR sequence to alter expression. In this case, reversing the sequence could still allow for the complementary binding of the resulting antisense lncRNA with the *DMD* 3' UTR. However, further investigation is needed to determine the biological significance of this region in myoblasts.

Methods

Size control constructs were made by amplifying the Lemaire A region or middle region of the *DMD* 3' UTR using the specific primers CGTTTAGGTACCAGGAAGTCTTTTCCACATGGC and CGTTTACTCGAGCCACTCAGCTGACAGTTCTCAAATG for the Lemaire A region, and the primers CGTTTAGGTACCCATTTGAGAACTGTCAGCTGAGTGG and CGTTTACTCGAGCTTAAACTTCTTAGTAGGATGTAAAGTAACCCCTTG to amplify the middle region of the 3' UTR. The full length *DMD* 3' UTR Renilla construct that we had previously made was used as a template to amplify the vector for the size control constructs using the primers CGTTTAGGTACCCATTTGAGAACTGTCAGCTGAGTGG and CGTTTACTCGAGCCTAGAATTACTGCTCGTTCTTCAGCAC for the Lemaire A insert and the primers CGTTTAGGTACCCAAGGGGTTACTTTACATCCTACTAAGAAGTTTAAG and CGTTTACTCGAGCCACTCAGCTGACAGTTCTCAAATG for the middle region insert. Overhang sequences in insert and vector primers contain the cut sites for the KpnI and XhoI restriction enzymes so that the inserted Lemaire A or middle region inserts in the reverse orientation in the *DMD* 3' UTR. The vector and insert PCR product was

digested with XhoI (New England Biolabs) and KpnI (New England Biolabs) and ligated using T4 DNA Ligase (New England Biolabs) as specified by the manufacturer. Transfections of C2C12 cells and the dual luciferase assay to measure relative expression levels of the reporter constructs were done as previously described (Chapter 1).

Mouse 3' UTR constructs

In our experiments, we analyzed the effects of a human *DMD* 3' UTR in the mouse C2C12 cell line. To determine if constructs containing the mouse *DMD* 3' UTR would give rise to different patterns of expression than the human *DMD* 3' UTR in C2C12 cells, we designed a Renilla reporter construct containing the mouse *DMD* 3' UTR downstream of the Renilla coding sequence in pHRL-CMV (3' UTR). This reporter construct (Mouse 3' UTR) along with the control Renilla construct (pHRL) and the original *DMD* 3' UTR construct (Human 3'UTR) was transfected into C2C12 cells, and protein expression levels were measured in myoblasts and myotubes using a dual luciferase assay (Figure C.3). Expression of the human *DMD* 3' UTR construct had a similar expression pattern as before (Chapter 1). We did not see a significant difference in expression levels with the pHRL construct containing the human or mouse *DMD* 3' UTR in C2C12 myoblasts or myotubes (Figure C.3), which is consistent with the high degree of conservation between the two 3' UTRs.

Methods

The full length mouse *DMD* 3' UTR was amplified from DNA extracted from C2C12 myoblasts using PCR and the following specific primers GCCGCTTTCTAGAAGGAAGCCTTTTCCACATGGC and

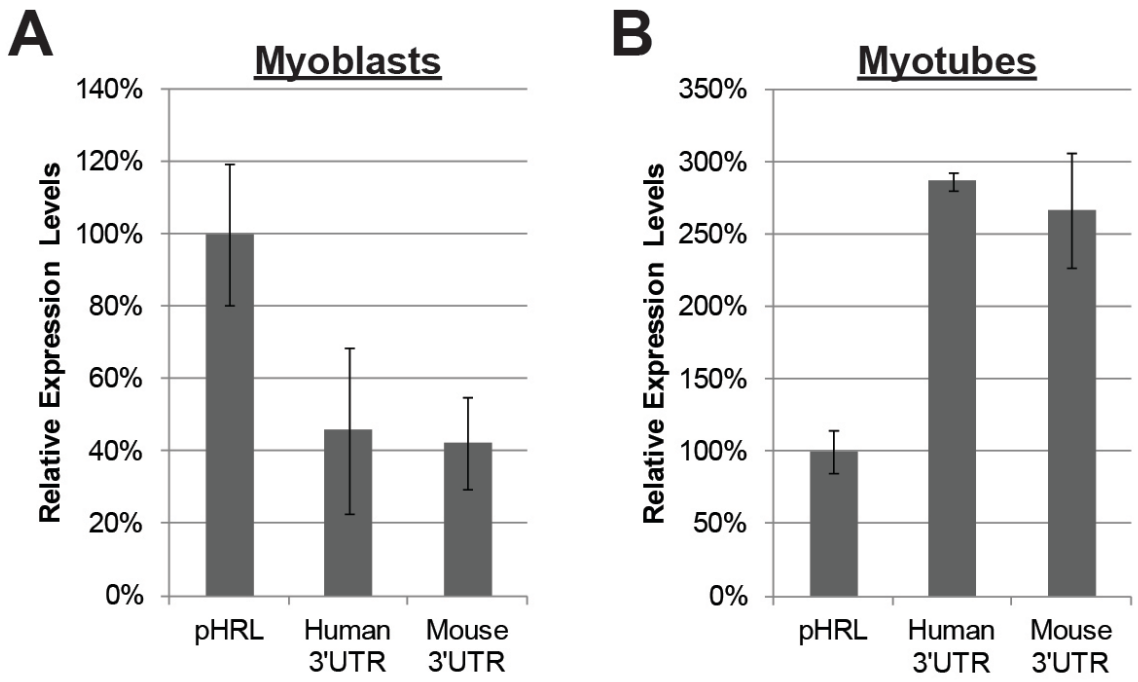


Figure C.3. Expression of the pHRL vector containing either the human or mouse *DMD* 3' UTR in C2C12 cells.

A. Relative protein expression of the control Renilla vector (pHRL), a construct containing the human *DMD* 3' UTR (Human 3'UTR), and a construct containing the mouse *DMD* 3' UTR (Mouse 3'UTR) in transfected C2C12 myoblasts are shown. B. Relative protein expression of the control pHRL construct, the human 3' UTR construct and the mouse 3' UTR construct in transfected C2C12 myotubes are shown. The result of three biological replicates is shown with expression levels normalized to pHRL-CMV expression (pHRL) in myoblasts or myotubes using a Renilla/Firefly dual luciferase assay. Error bars equal +/- 1 standard deviation.

GTCTCATGCCTAGGGGATCCGGCTAAGGCAGGATGGAACACT that include an overhang containing the XbaI and BamH1 restriction sites, respectively. The resulting PCR product was digested with XbaI and AvrII and ligated into the XbaI and BamHI restriction sites of pHRL-CMV (Promega) immediately downstream of the Renilla Luciferase ORF. The resulting plasmid contains the mouse *DMD* 3' UTR in place of the SV40 polyA region of the pHRL-CMV vector. Transfections of C2C12 cells and the dual luciferase assay to measure relative expression levels of the reporter constructs were

done as previously described (Chapter 1).

Optimization of C2C12 transfections

For our transfections in C2C12 cells, we optimized our transfection conditions by serially transfecting cells using differing amounts of transfected DNA, Lipofectamine 2000 (Life Technologies), and cell concentrations. We determined the optimal transfection contained $1x10^4$ C2C12 myoblasts with 25 ng of total transfected DNA 0.3uL of Lipofectamine 2000 reagent in a 96 well half area plate (Corning) with 0.16 cm^2 growth area. We saw no decrease in transfection efficiency when these amounts were scaled up for larger transfections. After incubating overnight at 37°C, the media was removed and replaced with growth medium consisting of Dulbecco's modified Eagle's medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Hyclone). Two days after transfection, the cells were harvested or placed in a low serum media (DMEM/F-12 (Life Technologies) supplemented with 2% horse serum (Sigma)) for 6 days to differentiate the cells before harvesting.

Cell death and variation decreased when the cells were spun down and resuspended in growth media after trypsinization and before transfection. We found a significant decrease in measured Firefly levels between wells that had a small amount of DMEM left in during lysis versus wells that were free of DMEM media which affected Renilla/Firefly ratios. Care should be taken to ensure that the DMEM media is completely removed from the cells prior to lysis and analysis. We found that freezing the cells for 30 minutes at -70°C and thawing helped in lysing myotubes, but did not see any benefit in multiple freeze thaw cycles. While measuring Renilla and Firefly luciferase levels, we found that light emitted from a well can bleed over into adjacent wells and

increase the measured luciferase levels in adjacent wells. For all transfections measuring protein levels in this dissertation, cells were transfected in every other well so that each well was surrounded by empty wells.

C2C12 cells comprise a heterogeneous group of cells, and not all cells differentiate into myotubes. We observed variability in different batches of C2C12 cells on how well they differentiated even from aliquots taken from the same stock. We also found that transfection conditions, such as amount of DNA and cell concentration, had an effect on how well the cells differentiated. Care should be taken when optimizing transfection conditions to not only optimize for increased expression of transfected constructs, but to also optimize differentiation, and every transfection should be monitored to ensure they are differentiating. In our attempt to increase expression of our transfected constructs in C2C12 myotubes, we transfected cells using the optimized conditions described in a previous publication comparing transfection methods in C2C12 cells (Dodds et al., 1998). Expression of our transfected constructs increased by ~10 fold using this described method, but our cells failed to differentiate even after 7 days in differentiation media. Low cell concentrations and high DNA amounts inhibited differentiation of C2C12 cells.

Despite the variability in C2C12 cells, we observed the same pattern of expression of our *DMD* 3' UTR deletion constructs under many different conditions. The transfections results presented in this dissertation are from the same batch of C2C12 cells transfected at the same time to ensure that we were accurately measuring changes in expression. However, the transfections were repeated several times in consecutive transfections of different batches of cells, in serial transfections of the same batch of

cells, under varying transfection conditions, and we obtained the same results except in cases where the cells visually did not form myotubes. All results presented in this dissertation were from C2C12 cells between passages 3 and 9. However, we transfected our deletion constructs in cells at passage 20 and obtained the same results. We also found that the increase in expression due to the *DMD* 3' UTR was dependent on the differentiation state of the C2C12 cells. 3' UTR reporter constructs still had an increase in expression in cells that were partially differentiated, but not to the degree seen when the cells were fully differentiated. In our hands, the cells were fully differentiated at Day 5 or 6 in differentiation media.

The experiments described in Appendix A were performed before we had optimized the conditions to maximize the differentiation of C2C12 cells. The protein expression of the cells described in Appendix A was at the levels of the C2C12 cells differentiated for 3 days described in Chapter 1, suggesting that the cells in Appendix A were not fully differentiated. However, the expression pattern was the same with an increase in expression during differentiation when the *DMD* 3' UTR was present and a large decrease when the Lemaire A and D regions were deleted, but the increased expression is greater when the cells are more differentiated. We conclude that the effect on stability shown in Appendix A is the effect in C2C12 cells differentiated for 3 days, but that we would have seen a larger effect on stability if these cells were more differentiated. All experiments in this dissertation besides the experiments described in Appendix A were done under conditions to optimize the differentiation of C2C12 cells.

References

Bovolenta, M., Erriquez, D., Valli, E., Brioschi, S., Scotton, C., Neri, M., Falzarano, M.S., Gherardi, S., Fabris, M., Rimessi, P.*, et al.* (2012). The DMD locus harbours multiple long non-coding RNAs which orchestrate and control transcription of muscle dystrophin mRNA isoforms. PLoS One *7*, e45328.

Dodds, E., Dunckley, M.G., Naujoks, K., Michaelis, U., and Dickson, G. (1998). Lipofection of cultured mouse muscle cells: a direct comparison of Lipofectamine and DOSPER. Gene Ther *5*, 542-551.

Hogg, J.R., and Goff, S.P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. Cell *143*, 379-389.

Tanguay, R.L., and Gallie, D.R. (1996). Translational efficiency is regulated by the length of the 3' untranslated region. Mol Cell Biol *16*, 146-156.

APPENDIX D

CXMD DOG SEQUENCES AND PRIMERS

The entire sequence of the LINE-1 insertion in intron 19 of the dystrophic Labrador retriever was determined using long range PCR (Figure D.1). The resulting pseudoexon insertion in the *DMD* mRNA along with the primers used to amplify the sequence is shown (Figure D.1). PCR primers were designed to amplify exons 3 through 32 and exons 44 through 50 from dog genomic DNA (Table D.1).

Figure D.1. DNA sequence of the LINE-1 insertion found in a dystrophic Labrador retriever.

The numbered uppercase sequence indicates the 6,363 nucleotide insertion found in intron 19. The lowercase sequence before and after the insertion indicate the surrounding intron 19 sequence at the insertion point. The underlined sequence indicates the resulting pseudoexon that is inserted into the *DMD* mRNA which includes 19 nucleotides of intron 19 and 165 nucleotides of the LINE-1 insertion. The boxed sequence shows the target site duplication at the beginning and end of the LINE-1 insertion. The forward and reverse primers used to amplify the insertion are shown in bold at the beginning and end of the intronic sequences surrounding the insertion.

Forward primer →

tgaaagtaagagctgagtcatggtacttcatttatgacctcatgtaaatatttattccttttgt
cctcaaaaattgagataaagacaggtcattcttatacagtgttatattatgcaaccaatacaaat
ttatagtttatgtatgttttattgtgctaacaatatttaacatatattcatagtgtaaaatatgt
ataagaatagcataattatgcttttaaaagcattttgttttctatgtatctctccatatatagct
gaattttctacttcagataatttactataagattagagtgctttttaaatatgtggaatgattt
tagaagtcctcttttaaatttgtcctcagaatgaaaatctaagacctg

```
    1 AGATGTGGAG CAAGATGGCG GAAGAGTAGG GTCCCCAAAT CACCTGTCTC
   51 CACCAAACTA CCTAGAAAAC CTTCAAATTA TCCTGAAAAT CTATGAATTC
  101 GGCCTGAGAT TTAAAGAGAG ACCAGCTGGA ATGCAACAGT GAGAAGAGTT
  151 CGCGCTTCTA TCAAGGTAGG AAGACGGGGA AAAAGAAGTA AAGGAACAAA
  201 GGCCTCCAAG GGGGAGGGGC CCCGCGAGGA CCCGGGCTGA GGCCGGGGCG
  251 AGTGTCCCCA GGACAGGAGA GCCCCGTCCC GGAGGAGCAG GAGCTGCACC
  301 GACCTTCCCG GGGGAAAGGG GCTCGCAGGG AGTTGGAGCA GGACCCAGGA
  351 GGGCGGGGAT GCCCTCGGGC TCCCTGGGAC AGTAACAGAG CAACTGCGCG
  401 CCCAGGAGAG TGCGCCGAGC TCCCTAAGGG CTGCAGCAGG CACGGCGGGA
  451 CCCGGCGGGA CCGGAGCAGC CAGGAGGGGC TCGGGCGGCG GCTCCGCGGA
  501 GGGGGCTGCG CGGCCCCGGG AGCAGCTCGG AGGGGCTCGG GCAGAGGAAG
  551 AGGCTCCGTG CGGAGGGGGC TGCGCGGTTC CAGGAGCAGC TCGCAGGGGC
  601 TCGGGCGGCG GCTCCGCGGA GGGGGCTGCG CGGCCCGGGA GCGCGAATCC
  651 ACCAGCGCAG GCTCCGGAGC ACAGGGCGCC GGGACACAGC CCAGGATCCC
  701 GCCTCCCCCG GGACAGGCAG AGGCCGGGAG GGCCCAGGAC AGCGAGGACG
  751 CTCCTGCCCC AGCTGAGCAG AGCAGCGGCC CCGCCCCGGA GCCTCCAGGC
  801 CCTGCAGACG GAGTTCCTGC CGGAGCTGAA TCCAGGTTTC CAGAGCTGCC
  851 CCGCCACTGG GGCTGTTCCT CCTGCGGCCT CACGGGGTAA ACAACCCCCA
  901 CTGAGCCCTG CACCAGGCAG GGGCACAGCA GCTCCCCCAA CTGCTAACAC
  951 CTGAAAATCA GCACAACAGG CCCCTCCCCC AGAAGATCAG CTAGACTGAC
 1001 AACTTCCAGG AGAAGCCAAG GGACTTAAAG AACACAGAAT CAGAAGATAC
 1051 TCCCCTGTGG TTCTTTTTTT TGTTTTGTTT TGTTTTTTTT TGTTTTTGTT
 1101 TTTGTTTTGT TTTGCTTTTT GATTTGTTTC CTTCCCCCAC CCCCTTTTTT
 1151 TCTCCTTTCT TTTTCTCTTT TTCTTCCTTT TTTTTTTTCT CTCGTTTTTC
 1201 TTTTCTTTTC TTCCCTTTTT TTTCTCTTTC TCTTTTCTTT CCTTCTTTCT
 1251 CTCCTCTCTT TTTCTCTTTT TCCCAATACA ATTTGCTTTT GGCCACTCTG
 1301 CACTGAGCAA AATGACTAGA AGGAAAACCT CACCTCAAAA GAAAGAATCA
 1351 GAAACAGTCC TCTCTCCCAC AGAGTTACAA AATCTGGATT ACAATTCAAT
 1401 GTCAGAAAGC CAATTCAGAA GCACTATTAT ACAGCTACTG GTGGCTCTAG
 1451 AAAAAAGTAT AAAGGACTCA AGAGACTTCA TGACTGCAGA ATTTAGAGCT
 1501 AATCAGGCAG AAATTAAAAA TCAATTGAAT GAGATGCAAT CCAAACTAGA
 1551 AGTCCTAACG ACGAGGGTTA ACGAGGTGGA AGAACGAGTG AGTGACCTAG
 1601 AAGACAAGTT GATAGCAAAG AGGGAAACTG AGGAAAAAAG AGACAAACAA
 1651 TTAAAAGACC ATGAAGATAG ATTAAGGGAA ATAAACGACA GCCTGAGGAA
 1701 GAAAAACCTA CGTTTAATTG GGGTTCCCGA GGGCGCCGAA AGGGACAGAG
 1751 GGCCAGAATA TGTATTTGAA CAAATTCTAG CTGAAAACTT TCCCAATCTG
 1801 GGAAGGGAAA CAGGCATTCA GATCCAGGAA ATAGAGAGAT CCCCCCCTAA
 1851 AATCAATAAA AACCGTTCAA CACCTCGACA TTTAATTGTG AAGCTTGCAA
 1901 ATTCCAAAGA TAAGGAGAAG ATCCTTAAAG CAGCAAGAGA CAAGAAATCC
 1951 CTGACTTTTA TGGGGAGGAG TATTAGGGTA ACAGCAGACC TCTCCACAGA
 2001 GACCTGGCAG GCCAGAAAGG GCTGGCAGGA TATATTCAGG GTCCTAAATG
 2051 AGAAGAACAT GCAACCAAGA ATACTTTATC CAGCAAGGCT CTCATTCAAA
```

Figure D.1 continued

```
2101 ATGGAAGGAG AGATAAAGAG CTTCCAAGAC AGGCAGCAAC TAAAAGAATA
2151 TGTGACCTCC AAACCAGCTC TGCAAGAAAT TTTAAGGGGG CCTCTTAAAA
2201 TTCCCCTTTA AGAAGAAGTT CAGTGGAACA GTCCACAAAA ACAAAGACTG
2251 AATAGATATC ATGATGACAC TAAACTCATA TCTCTCAATA GTAACTCTGA
2301 ATGTGAACGG GCTTAATGAC CCCATCAAAA GGCGCAGGGT GTCTGACTGG
2351 ATAAAAAAGC AGGACCCATC TATTTGCTGT CTACAAGAGA CTCATTTTAG
2401 ACAGAAGGAC ACCTACAGCC TGAAAATAAA AGGTTGGAGA ACCATTTACC
2451 ATTCGAATGG TCCTCAAAAG AAAGCAGGGG TAGCCATCCT TATATCAGAT
2501 AAACTAAAAT TTACCCCAAA GACTGTAGTG AGAGATGAAG AGGGACACTA
2551 TATCATACTT AAAGGATCTA TTCAACAAGA GGACTTAACA ATCCTCAATA
2601 TATATGCTCC GAATGTGGGA GCTGCCAAAT ATATAAATCA ATTATTAACC
2651 AAAGTGAAGA AATACTTAGA TAATAATACA CTTATACTTG GTGACTTCAA
2701 TCTAGCTCTT TCTATACTCG ACAGGTCTTC TAAGCAAAAC ATCTCCAAAG
2751 AAACGAGAGC TTTAAATGAT ACACTGGACC AGATGGATTT CACAGATATC
2801 TACAGAACTT TACATCCAAA CTCAACCGAA TACACATTCT TCTCAAGCGC
2851 ACATGGAACT TTCTCCAGAA TAGACCACAT ATTGGGTCAC AAATCGGGTC
2901 TGAACCGATA CCAAAAGATT GGGATTGTCC CCTGCATATT CTCGGACCAT
2951 AATGCCTTGA AATTAGAACT AAATCACAAC AAGAAGTTTG GAAGGACCTC
3001 AAACACATGG AGGTTAAGGA CCATCCTGCT AAAAGATAAA AGGGTCAACC
3051 AGGAAATTAA GGAAGAATTA AAAAGATTCA TGGAAACTAA TGAGAATGAA
3101 GATACAACCG TTCAAAATCT TTGGGATGCA GCAAAAGCAG TCCTAAGGGG
3151 GAAATACATC GCAATACAAG CATCCATTCA AAAACTGGAA AGAACTCAAA
3201 TACAAAAGCT AACCTTACAC ATAAAGGAGC TAGAGAAAAA ACAGCAAATA
3251 GATCCTACAC CCAAGAGAAG AAGGGAGTTA ATAAAGATTC GAGCAGAACT
3301 CAATGAAATC GAGACCAGAA GAACTGTGGA ACAGATCAAC AGAACCAGGA
3351 GTTGGTTCTT TGAAAGAATT AATAAGATAG ATAAACCATT AGCCAGCCTT
3401 ATTAAAAAGA AGAGAGAGAA GACTCAAATT AATAAAATCA TGAATGAGAA
3451 AGGAGAGATC ACTACCAACA CCAAGGAAAT ACAAACGATT TTAAAAACAT
3501 ATTATGAACA GCTATACGCC AATAAATTAG GCAATCTAGA AGAAATGGAC
3551 GCATTCCTGG AAAGCCACAA ACTACCAAAA CTGGAACAGG AAGAAATAGA
3601 AAACCTGAAC AGGCCAATAA CCAGGGAGGA AATTGAAGCA GTCATCAAAA
3651 ACCTCCCAAG ACACAAGAGT CCAGGGCCAG ATGGCTTCCC AGGAGAATTT
3701 TATCAAACGT TTAAAGAAGA AATCATACCT ATTCTCCTAA AGCTGTTTGG
3751 AAAGATAGAA AGAGATGGAG TACTTCCAAA TTCGTTCTAT GAAGCCAGCA
3801 TCACCTTAAT TCCAAAGCCA GACAAAGACC CCGCCAAAAA GGAGAATTAC
3851 AGACCAATAT CCCTGATGAA CATGGATGCA AAAATTCTCA ACAAGATACT
3901 GGCCAATAGG ATCCAACAGT ACATTAAGAA AATTATTCAC CATGACCAAG
3951 TAGGATTTAT CCCTGGGACA CAAGGCTGGT TCAACACCCG TAAAACAATC
4001 AATGTGATTC ATCATATCAG CAAGAGAAAA ACCAAGAACC ATATGATCCT
4051 CTCATTGGAT GCAGAGAAAG CATTTGACAA AATACAGCAT CCATTCCTGA
4101 TCAAAACTCT TCAGAGTGTA GGGATAGAGG GAACATTCCT CGACATCTTA
4151 AAAGCCATCT ACGAAAAGCC CACAGCAAAT ATCATTCTCA ATGGGGAAGC
4201 ACTGGGAGCC TTTCCCCTAA GATCAGGAAC AAGACAGGGA TGTCCACTCT
4251 CACCACTGCT ATTCAACATA GTACTGGAAG TCCTAGCCTC AGCAATCAGA
4301 CAACAAAAAG ACATTAAAGG CATTCAAATT GGCAAAGAAG AAGTCAAACT
4351 CTCCCTCTTC GCCGATGACA TGATACTCTA CATAGAAAAC CCAAAAGTCT
4401 CCACCCCAAG ATTGCTAGAA CTCATACAGC AATTCGGTAG CGTGGCAGGA
4451 TACAAAATCA ATGCCCAGAA GTCAGTGGCA TTTCTATACA CTAACAATGA
4501 GACTGAAGAA AGAGAAATTA AGGAGTCAAT CCCATTTACA ATTGCACCCA
4551 AAAGCATAAG ATACCTAGGA ATAAACCTCA CCAAAGATGT AAAGGATCTA
```

Figure D.1 continued

```
4601 TACCCTCAAA ACTATAGAAC ACTTCTGAAA GAAATTGAGG AAGACACAAA
4651 GAGATGGAAA AATATTCCAT GCTCATGGAT TGGCAGAATT AATATTGTGA
4701 AAATGTCAAT GTTACCCAGG GCAATATACA CGTTTAATGC AATCCCTATC
4751 AAAATACCAT GGACTTTCTT CAGAGAGTTA GAACAAATTA TTTTAAGATT
4801 TGTGTGGAAT CAGAAAAGAC CCCGAATAGC CAGGGGAATT TTAAAAAAGA
4851 AAACCATATC TGGGGGCATC ACAATGCCAG ATTTCAGGTT GTACTACAAA
4901 GCTGTGGTCA TCAAGACAGT GTGGTACTGG CACAAAAACA GACACATAGA
4951 TCAGTGGAAC AGAATAGAGA ATCCAGAAGT GGACCCTGAA CTTTATGGGC
5001 AACTAATATT CGATAAAGGA GGAAAGACTA TCCATTGGAA GAAAGACAGT
5051 CTCTTCAATA AATGGTGCTG GGAAAATTGG ACATCCACAT GCAGAAGAAT
5101 GAAACTAGAC CACTCTCTTT CACCATACAC AAAGATAAAC TCAAAATGGA
5151 TGAAAGATCT AAATGTGAGA CAAGATTCCA TCAAAATCCT AGAGAAGAAC
5201 ACAGGCAACA CCCTTTTTGA ACTCGGCCAT AGTAACTTCT TGCAAGATAC
5251 ATCCACAAAG GCAAAAGAAA CAAAAGCAAA AATGAACTAT TGGGACTTCA
5301 TCAAGATAAG AAGCTTTTGC ACAGCAAAGG ATACAGTCAA CAAAACTCAA
5351 AGACAACCTA CAGAATGGGA GAAGATATTT GCAAATGACA TATCAGATAA
5401 AGGGCTAGTT TCCAAGATCT ATAAAGAACT TATTAAACTC AACACCAAAG
5451 AAACAAACAA TCCAATCATG AAATGGGCAA AAGACATGAA CAGAAATCTC
5501 ACAGAGGAAG ACATAGACAT GGCCAACATG CATATGAGAA AATGCTCTGC
5551 ATCACTTGCC ATCAGGGAAA TACAAATGAA AACTACAATG AGATACCACC
5601 TCACACCAGT GAGAATGGGG AAAATTAACA AGGCAGGAAA CAACAAATGT
5651 TGGAGAGGAT GCGGAGAAAA GGGAACCCTC TTACACTGTT GGTGGGAATG
5701 TGAACTGGTG CAGCCACTCT GGAAAACTGT GTGGAGGTTC CTCAAACAGT
5751 TAAAAATATA CCTGCCCTAC GACCCAGCAA TTGCACTGTT GGGGATTTAC
5801 CCCAAAGATA CAAATGCAAT GAAACGCCGG GACACCTGCA CCCCGATGTT
5851 TCTAGCAGCA ATGGCCACGA TAGCCAAACT GTGGAAGGAG CCTCGGTGTC
5901 CAACGAAAGA TGAATGGATA AAGAAGATGT GGTTTATGTA TACAATGGAA
5951 TATTACTCAG CTATTAGAAA TGACAAATAC CCACCATTTG CTTCAACGTG
6001 GATGGAACTG GAGGGTATTA TGCTGAGTGA AGTAAGTCAG TCAGAGAAGG
6051 ACAAACATTA TATGTTCTCA TTCATTTGGG GAATATAAAT AATAGTGAAA
6101 GGGAAAATAA GGGAAGGGAG AAGAAATGTG TGGGAAATAT CAGAAAGGGA
6151 GACAGAACAT AAAGACTGCT AACTCTGGGA AACGAACTAG GGGTGGTAGA
6201 AGGGGAGGAG GGCGGGGGGT GGGAGTGAAT GGGTGACGGG CACTGGGGGT
6251 TATTCTGTAT GTTAGTAAAT TGAACACCAA TAAAAAAATA AATTAAAAAA
6301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
6351 AAAAAAAAGA AAA
```

tctaagacctgccaaaaaaaattcttcaatacttaagttttataccagttttgaaccagagtaat
acatcaaatatat***tttcaattcacttttggcga***

← Reverse Primer

Table D.1

Primers used to amplify *DMD* exons in dogs.

Exon	Forward Primer	Reverse Primer
3	TGCCCTTCCTAGAATCAACA	GTCTCCCATCCTGTAGGTCA
4	GGATCCACAAGAGTTCATGC	CACAGAGTGCCCTATCAGGT
5	TTGTTTTCTTCCCAAATGGA	TCCATCCCATCTCAAATGAC
6	GTGGTTCTTGTTCAGGGAAG	TAGCAGGAAATGAGCTGGAA
7	TTAAGGAATGATGGGCATGG	TGCATAGTTTCTCTTTCATGCTT
8	TTGTTCATTCGCCTTTCAGA	TGTGTATGATGGCTAAAAATGC
8 alt*	TGCCAAATTGGAAGGACATT	CATTTCCACTTCCTGGATGG
9	GAAAGCAGTCTTTCCAGGCTTA	GCTCCTGATGAAGGAGGAAA
10	TGTAATGTACTGGAACCATCTGC	ATGACAAGTCATGGGCATTG
11	TCCGTAGGGATTGATCTAAAGTT	GATTCCAAAACTCCCCATTTT
12	GGGCTCTGCTTCCCTTACTT	CATGTACGCATGTGTGTTTGT
13	TTTGGGAAGGAATGTTTTGG	GAAACGTCTGCAAATGGGATA
14	GGTGTGACAACTGCCTCAGA	CAATTTCCCCCACATGACTC
15	TGGCAGCGTTTTACTGAAGA	AGCAAGAACAATGAACATTTGG
16	GATCCTTTGTGGGCATGATT	CCCATTCAGTTTCCCTCCTT
17	TATTTCCCTTTGCCACTCCA	TGAGTTTTCCCGCTTAACAG
18	ACAGTAACAGGGGCATCAGG	GCACAGAGTTTACAAACAGCAAA
19	TTAAGGCTTGAAAGGGCAAA	GCACATCCCATTTTCTTCCA
20	CAGTCTGTGGCTTCAGTGGA	TCCAAATAGGAGGAAAGAACCTT
21	GAGGGTGGCCTAAACTGTCA	TCAGGTTAGCCATTTCAAGGA
22	GCAGTGTGATTTAGTTTGCTGA	GGAGGGAAGACACAAATCCA
23	TTGAATCACATAGCTTTCCCTGT	CGCATCATTAGGGGAGAAAC
24	TGGTTCCTTGGCCTATGTTC	CAGGGGAGCAAAATCACACT
25	CAATTTTTGCAAGGTTGAGACA	AAATGGAGACAAGGCACAGC
26	TCATGCTTTGCTTGTTGGTT	TCCACCATTAAACCAAGGAAAT
27	AGGGATCTTGTGGGGAAGAA	CAGCAGTGTTTCTGCTCTGA
28	ACTGCAAAGGAAAGGTGAGC	AGTGTCAAGGAAGGCAAAGC
29	TTTCCTGTTTTATCACCCCAAT	GGCCTAAAAGCATTTGCTAGA
29 alt*	TTTGATGCAACATTCAGAGGA	TTAAAACCTGGAAGCCAAGC
30	GCAGGATGGCAGGAAAAGTA	CCAGAGCTCTTTTCCCCTCT
31	TCCTGCAGAACATGCTGACT	TTGGGAGAATTCGGTAGCTG
32	AGCTACCGAATTCTCCCAAA	GCTGTGTATTTGCTTCCAGAAA
44	GCTATGTCCATGTGATTTGC	TCCATCATCCTTCACAACCT
45	AAGGACGTAGGGCTTCATTT	TCCTTAATCTTGGTGCCTTTC
46	TTTGGAGATATTGAAATTGCTTTG	ACCTAATGGGCAGGAAATCG
47	TGAGGTACCTGTTGGCCTTC	CCCTCAATCCCAGCAATAAA
48	AGCATAGGATTTGGGGGAAG	TGATATTGCCATTCTTTTGGAA
49	CCCAGGAAATTGAAGTAGCA	AGCACACCAGGACAGAATTG
50	TTGTAGGGTGGTTGGCTAAAA	CCAGGGATCACCTCAGTGTT

*Alternate nonoverlapping primer sets were designed for exons 8 and 29 to verify the deletion boundary in the Tibetan terrier containing a deletion of exons 8 through 29 of the *DMD* gene.